REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
30-06-2006	Final	Dec 2001 - Mar 2004

4. TITLE AND SUBTITLE
New Tools and Methods for Assessing Risk-Management Strategies

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00014-02-1-0179

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Terry P. J. F. Vendlinski, Allen Munro, Gregory K. W. K. Chung, Girlie C. Delacruz, Quentin A. Pizzini, William L. Bewley, Gale Stuart, and Eva L. Baker
CRESST/University of California, Los Angeles
BTL/University of Southern California

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UCLA CSE/CRESST
300 Charles E. Young Dr. North
300 GSE&IS/Mailbox 951522
Los Angeles, CA 90095-1522

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report describes research conducted by the UCLA Center for Research on Evaluation, Standards, and Student Testing (CRESST) and its subcontractor the University of Southern California Behavioral Technology Laboratories (BTL) on two tools designed to help evaluate the process used by novice acquisition officers to integrate risk-management strategies with the federal military acquisition process. The Human Performance Knowledge Mapping Tool (HPKMT) was used to evaluate understanding of the key phases in the acquisition process and of risk management. The Decision Analysis Tool (DAT) allowed students to use Expected Value and Multi-attribute Utility Theories to evaluate the risks and benefits of various acquisition alternatives and allowed researchers to monitor the process students used to make a procurement decision. Comparison of the HPKMT knowledge maps of 17 students against expert maps indicated that student understanding of use of risk management in the acquisition process trended higher during training but did not improve significantly. Sequential analysis of data from the DAT revealed distinct risk-management strategies.

15. SUBJECT TERMS
risk managements, acquisition process, expected value theory, multi-attribute utility theory, HPKMT, DAT, decision analysis

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*
Unclassified	Unclassified	Unclassified			

New Tools and Methods for Assessing
Risk-Management Strategies

Deliverable – March 2004

Knowledge, Models, and Tools to Improve the
Effectiveness of Naval Distance Learning

Eva L. Baker
National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
University of California, Los Angeles

Office of Naval Research
Award # N000140210179

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
301 GSE&IS, Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under Office of Naval Research Award Number #N00014-02-1-0179, as administered by the Office of Naval Research. The findings and opinions expressed in this report do not necessarily reflect the positions or policies of the Office of Naval Research.

TABLE OF CONTENTS

NEW TOOLS AND METHODS FOR ASSESSING RISK-MANAGEMENT STRATEGIES [1]

Terry P. J. F. Vendlinski[a], Allen Munro[b], Gregory K. W. K. Chung[a], Girlie C. Delacruz[a], Quentin A. Pizzini[b], William L. Bewley[a], Gale Stuart[a], and Eva L. Baker[a]

[a]CRESST/University of California, Los Angeles

[b]BTL/University of Southern California

Abstract

At the request of the Office of Naval Research (ONR), we provided the U.S. Navy two tools to help evaluate the process that novice acquisition officers use to integrate risk-management strategies with the federal military acquisition process. The Human Performance Knowledge Mapping Tool (HPKMT) was designed to evaluate the ability of subjects to depict the relationships among the key phases in the acquisition process and subject understanding of risk management. The Decision Analysis Tool (DAT) allowed subjects to use Expected Value and Multi-attribute Utility Theories to evaluate the risks and benefits of various acquisition alternatives, and allowed us to monitor the process subjects used to arrive at a procurement decision. When we evaluated the HPKMT knowledge maps of 17 subjects against expert maps developed by their instructors, we found that subject understanding of incorporating risk management in the acquisition process trended higher, but did not improve significantly. Sequential analysis of data from the DAT allowed us to isolate distinct risk-management strategies, as well as strategies that overly focused on (or ignored) aspects of risk management. The use of a referent to determine the conceptual relationships and strategic acquisition skills necessary to be a skilled acquisition officer are discussed as extensions of this work.

---

[1] We would like to thank Dr. Mary Lantgen, CAPT Frank Camelio, CAPT Ronald Crowell, CDR John Hill, CDR (S) Scott Heller, CDR Michael Giaque, LCMDR Robert Phillips, and the students who participated in this study. We also wish to thank Dr. Zenaida Aguirre-Muñoz who assisted with early parts of this study, and Ms. Joanne Michiuye of UCLA/CRESST for her help with the preparation of this manuscript and with data collection.

Introduction

Not all tests and assessments support equally well the different types of inferences we make about student learning and how students solve problems. Assessments that only allow us to view a student's final answer, for example, inherently limit the inferences we can make about that student's ability in the domain of interest. Often this limitation means we can only reasonably infer that a student *can* answer the question posed. Repeated administrations of similar tests may also allow us to conclude with some certainty that the student can *routinely* answer the question posed. However, the data generated by such assessments often do not allow us to infer with any degree of certainty that a student arrived at their answer using the process we expected or using the tools we desired. Overgeneralizing the results from a single type of assessment, especially in instructional settings, has been a perennial problem and one that makers of tests caution their users against (AERA, APA, & NCME, 1999; Baker, Linn, Herman, & Koretz, 2002; Crouse & Trusheim, 1988; Educational Testing Service, n.d.). Early in the last century, Vygotsky (Cole, John-Steiner, Scribner, & Souberman, 1978) argued that our inferences about student understanding must be bounded by the amount of support a student received to achieve that answer. In recent times, it seems logical to include the tools at a student's disposal during the administration of the test in addition to the people and cultural artifacts with which the student interacts (Norman, 1993). Others have argued that we should also consider the uses and consequences of the inferences we will make from assessment data when designing assessment modalities (Messick, 1989; Popham, 2000). For example, the test required to conclude someone has memorized a list of emergency procedures would probably look different from the test required to infer that someone knew why each of those procedures was on the list in their prescribed order and how to accomplish each procedure.

The Need to Consider the Context of an Assessment

We must consider the context in which the student posited an answer in order to accurately infer a student's understanding of a body of knowledge (Linn, Dunbar, Harnish, & Hastings, 1982). For example, a student who is asked to choose the correct answer from a list of possible answers might not have the depth of understanding of a student who is asked to create an answer without the assistance

of a list. Obviously, the quality of the list of alternatives impacts our inferences of student understanding so it is also an overgeneralization to argue that selected-response tests only allow us to measure student recall or basic knowledge. The inferences made about students choosing the correct answer from a list of answers that are created using common misconceptions are qualitatively different from the inferences made about students identifying the correct answer from a list of "unattractive" distracters.

The Need to Consider How We Will Use Assessment Data

It seems we must also consider how we will use our inferences. A test that requires test takers to pick an answer from a list of answers may provide little data to support inferences that students could or would actually generate the answer they chose from the list on their own. Optimally, inferences about a student's ability to generate an answer should be based on an evaluation of what happens when the student attempts to solve a problem or is asked to self-generate an answer. Selecting an answer from a list would only allow one to conclude a student can pick the answer, but the requirement to generate an answer would allow one to conclude a student might be able to arrive at an answer in a less supported context and therefore that the latter student understands more deeply than the former. Generation tasks, however, should not be viewed as absolutely "better" than selection tasks since we might only need to confirm a student can "pick" a solution rather than "create" a solution. For example, the ability to distinguish a red light from a green light or to read letters on an eye chart as part of a driving test is an appropriate real-world "selection" task. In these cases, we are not interested in the "how" a person answers the question, only that they are able to answer correctly. On the other hand, when we want to infer a student has the ability to create a solution or to perform a task, such selected-response assessments probably do not provide us with the data we require to make necessary evaluations, unless the correlations between an ability to select and an ability to apply a concept have been demonstrated. When determining if people can actually drive a car safely, for example, it is probably more appropriate to evaluate their driving rather than give them a pencil-and-paper test about driving (Wiggins, 1993).

The Challenge of Timeliness of Inferences

Unfortunately, the time it takes for students to complete and for instructors to evaluate the open-ended process of producing an answer is often much greater than the time required to evaluate a selected-response answer as either correct or incorrect. As the number of steps from problem statement to problem solution increases, the time necessary to generate and evaluate the transitions between each step also increases. As solutions become more complex (involving many concepts) or if the steps to a solution can be ordered in many ways, the time required to make accurate evaluations of the ability of a test taker is further increased.

Modern computer technology offers the opportunity to overcome many of the above limitations. First, we can design computerized interfaces that allow us to both offer test takers an almost unlimited problem-solving space and to record every interaction they make within the problem space. These interfaces also allow us to control the external tools and artifacts test takers have access to when they attempt to solve the problem. Second, tools based on fast computational models provide the means necessary to organize and analyze the voluminous amounts of data that often result from computerized assessment tools and minimize instructor evaluation time. We developed tools and methods to aid instruction in a course for the U.S. Navy based on course content, context, and the anticipated uses of assessment data.

Engineering Duty Officer School

The Office of Naval Research (ONR), as part of its Capable Manpower Future Naval Capability (FNC) program, has funded the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) to develop assessment models and tools for use in Navy and Marine operational environments. One of the environments selected by ONR was the Engineering Duty Officer (EDO) School. Specifically we created models and automated tools to help evaluate the process EDO students used to develop their frigate acquisition plan and resolve a particular acquisition problem.

The Navy's Engineering Duty Officers (EDOs) manage large-scale development and procurement processes. During their initial training in a six-week EDO Basic Course at the EDO School at Port Hueneme, California, EDO candidates are taught about making complex decisions as part of project risk management. The students, who have higher degrees in one or more engineering disciplines, must

learn to make complex decisions that incorporate the uncertainty of future events, and to convincingly present their acquisition recommendations to senior Navy officers for approval. During the Basic Course, students are given a variety of techniques for mitigating project risk and for making complex decisions. Exercises are conducted in which three teams of approximately six students each analyze risks in assigned projects and make formal presentations to boards of reviewing officers, who are, in this case, faculty members of the EDO School.

While the briefing process allows school instructors some insight into the process students use to resolve a system acquisition problem, this insight is necessarily "filtered" through the students because it is the students who select the aspects of their solution they think important to present. In other words, the instructors have to rely on what the students present and their own experience with previous students, rather than more detailed empirical evidence, to evaluate the proposed solution to the problem described in each presentation. This method can make the evaluation of the processes students used to reach their conclusions difficult.

When we approached the EDO School faculty about their needs for effective training and assessment of students, the faculty members asked for help in training and assessing decision-making skills in the context of the assigned project exercises. In particular, during the final exercise, students are asked to address a mid-procurement project crisis—the vendor of an important ship system (the Refueling at Sea system, or RAS) has decided to discontinue providing that system. Student teams must determine and evaluate possible solutions and present their recommendations to the review board. What could be done to make this experience one that reinforced decision-making skills taught in the course, and how could the students' application of those skills be assessed?

In addition to assessing student knowledge of risk-management and the federal military acquisition process, several related topics presented to the students in the course drew themselves to our attention. One of these was the topic of multi-attribute measures of utility. Early in the course, students are presented with an example of choosing a restaurant for dinner. Four possible restaurants are considered, and each is given a simple numerical score on such attributes as nearness, expense, atmosphere, and food quality. The concept of weighting utility scores differentially is also introduced, and a simple Excel worksheet for computing

the "best" restaurant outcome based on weighted attribute scores is presented. At this stage of the course, the students have been exposed to these concepts:

- Multiple components of utility (attributes)
- Weighted attribute values
- Use of computer-based tools to support decision modeling

Later during the course, the faculty briefly introduces Expected Value Theory as a more sophisticated framework for making such complex decisions. In addition to estimating an outcome value for each alternative choice (by summing the weighted utility values of all the potential consequences of that choice), the students also make estimates of the probability of each outcome, given the preceding choice. The expected value of a decision is computed by summing the probability-weighted estimated outcomes of each decision. Although it would be possible to repeatedly make such estimates of probability and utility values and to repeatedly compute expected values by hand, this task clearly would benefit from the use of computer-based support tools. At this stage of the course, the students have also been exposed to these concepts:

- Alternative decision outcomes can be assigned estimated probabilities of occurrence.
- Expected values can be computed from estimated probabilities and the sums of weighted outcome utilities.
- Decisions can be made based on expected value analyses.

We decided to build an experimental tool based on the six concepts just described. The tool would have two purposes: to contribute to the students' understandings of these topics, and to provide a natural, problem-centered task for collecting data for assessment. The tool would have to be simple, so that students could learn to use it very quickly. That would make it possible for them to apply the tool to the RAS problem during the last project exercise of the course. In addition, it was decided that the tool would be delivered to the school with appropriate content to facilitate its use in the course. This content would include a simple version of the tool applied to the restaurant decision example, and would make it possible to introduce a simplified form of the tool when multi-attribute utility concepts were introduced early in the course. Second, a simple example—selecting a digital

camera—was developed for use in introducing the concept of expected value. Third, the RAS system decision would be implemented in the tool, so that students would be able to focus on the estimates they had to make, rather than on the mechanics of authoring every aspect of the alternatives from scratch using the tool.

Method

The National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at the University of California, Los Angeles (UCLA), and the Behavioral Technology Laboratories (BTL) at the University of Southern California (USC) supplied the Navy with two computer-based tools to generate the data necessary to evaluate the RAS acquisition process developed by students at the EDO school: the Human Performance Knowledge Mapping Tool (HPKMT) and the Decision Analysis Tool (DAT). In addition, to better understand the impact of these tools on their users, we asked each student to complete a short student survey at the end of the course.

Human Performance Knowledge Mapping Tool

The first tool we provided the Navy is known as the Human Performance Knowledge Mapping Tool (HPKMT). The HPKMT is designed to evaluate the ability of students to depict the relationships among a set of concepts. In the case of the EDO school, we provided EDO students a list of the phases in the federal military acquisition process and a list of the possible relationships between these processes, and then asked the students to place them in the appropriate order. The school's three instructors developed the list of phases in the acquisition process, the list of relationships between the phases, and the acceptable solution or "expert map" we used to evaluate student solutions.

One objective of this research was to test a method of capturing students' understanding of the problem-solving procedure related to risk management. A knowledge map is a graphical network representation of a person's understanding of a particular content area or process. CRESST has used knowledge maps for assessment purposes (e.g., to measure a student's content knowledge) across a variety of content areas and ages (e.g., Chung, Harmon, & Baker, 2001; Chung, O'Neil, & Herl, 1999; Herl, Baker, & Niemi, 1996; Herl, O'Neil, Chung, & Schacter, 1999; Klein, Chung, Osmundson, Herl, & O'Neil, 2002; Osmundson, Chung, Herl, & Klein, 1999). For the purposes of assessing problem solving, we used a

representation that conceptualized the network as consisting of the steps in the process and the transitions between steps.

Because we were attempting to develop a new format to measure process-related knowledge, we pilot tested two different formats. Our initial approach was to use links to denote transitions and the link labels to represent the reason behind the transition. This representation is similar to a flow chart with an important difference: The transitions are labeled and the labels were intended to represent the student's justification or reason for going from one step to another step. In one condition, we provided one group of students with a set of predefined links to select from, and an option to type in reasons of their own. One of the link options was an asterisk ("*"), which was provided to students so that they could specify a transition without providing an explicit reason—e.g., when the transitions are naturally occurring and the requirement to specify a reason would be awkward. The task for this condition is given in Appendix A. In the second condition, we provided students only with the asterisk and with the option for students to type in reasons of their own. The task for this condition is given in Appendix B.

For the pilot test, the list of steps were: *analyze idea, consider impacts of 15 ships + parts, consider impacts of configuration management, consider impacts of contracting, consider impacts of late Phase A, consider impacts of manufacturing, consider impacts of R & D, consider impacts of supportability, consider impacts of test and evaluation, generate idea, prioritize cost driver, prioritize decision drivers, prioritize performance drivers, prioritize schedule drivers, rate alternate ideas against environment with decision drivers,* and *select option.* In the condition where links were provided, the links were: *, *decide, define the cost environment, define the environment, define the performance environment, define the schedule environment*, and *score impacts.*

Based on results from the pilot test, several changes were made for the main study. First, the content of the knowledge-mapping task was modified to reflect a higher level of abstraction. Instead of focusing on a particular problem, there was more interest in measuring knowledge of the problem-solving process at a general level. Second, we eliminated the type-in option because very few students used it. Third, the requirement to specify a reason was eliminated due to apparent interference and confusion, and fourth, for logistical reasons, the task was administered on paper. The objective of the mapping task was to develop and test a method of assessing students' understanding of the process of analyzing risk within a program management context.

The task for the main study is given in Appendix C. The final set of steps for the main study were: *conduct analysis of alternatives, define cost constraints, define legal constraints, define performance constraints, define requirement to address a risk, define schedule constraints, define the environment, detect new risks, execute best solution, methodically determine alternatives, methodically pick best solution,* and *prioritize known risks.*

A pretest knowledge map was administered to students during the beginning of the course (Week 1). Students used the Decision Analysis Tool (discussed in the next session) near the end of the course for the RAS problem (Week 9). A posttest knowledge map was administered during the last week of the course (Week 10). In addition, a student survey was administered at Week 10, immediately after the students completed their knowledge map.

Analysis of Knowledge Map Data

Scoring of the student knowledge maps was done by comparing each student's map against a consensus criterion map developed by the instructors. Appendix D shows the criterion map used in the pilot study and Appendix E shows the criterion map used in the main study. The essential measure is the number of propositions (i.e., node-link-node) in the student map that are also in the referent map. Prior research has shown that in general, scoring student knowledge maps using expert-based referents has been found to discriminate between experts and novices (Herl, 1995; Herl et al., 1996), discriminate between different levels of student performance (Herl, 1995; Herl et al., 1996), relate moderately to external measures (Aguirre-Muñoz, 2000; Herl, 1995; Herl et al., 1996; Klein et al., 2002; Lee, 2000; Osmundson et al., 1999), detect changes in learning (Chung et al., 2001; Osmundson et al., 1999; Schacter, Herl, Chung, Dennis, & O'Neil, 1999), and be sensitive to language proficiency (Aguirre-Muñoz, 2000; Lee, 2000).

Decision Analysis Tool

The second tool we provided was developed as a joint project of CRESST and BTL. The technologies used to implement the Decision Analysis Tool (DAT) were VIVIDS (Munro & Pizzini, 1998; Munro, Surmon, Johnson, Pizzini, & Walker, 1999; Munro, 2003) and *iRides Author* (Munro, Surmon, & Pizzini, in press). This tool was designed to enable training developers to create interactive graphical simulations and training in the context of those simulations. The iRides program can deliver the

training specifications as a Java application, or over the Web as an applet or a Web Start application. The behavior specification language of iRides is sufficiently expressive and powerful that it was possible to create implementations of a real software tool for aiding decision making using weighted attributes and expected value theory.

The tool was developed in three phases, which resulted in three releases of the DAT: prototype, version 1, and version 2. After each of the first two phases, student usage and instructor comments led to significant revisions that appeared in the subsequent release. Some of these modifications were designed to make elements of the user interface easier to learn and to use, to correct algorithmic errors, and to improve data reporting. In addition, however, a number of changes were made to the tool to bring it into compliance with the specific content and within the context of the EDO Basic Course. Examples of this included restricting attribute utility values to integers between 1 and 5, and including three standard attributes of outcome utility: cost, performance, and schedule.

Using the DAT. In this discussion, the behavior of version 2.0 of the DAT is described. The data collection took place using version 1.0, but the major differences in 2.0 are not relevant to the core issues of operation sequencing in the usage of the tool. The primary difference in version 2.0 is that users are not limited in the depth and breadth of the decision trees that can be authored. In addition, the graphical user interface of 2.0 is improved by the use of Java Swing interface objects (sliders, radio button groups, check box groups, and the like) in place of authored iRides simulation objects with similar functionality.

If an author begins to develop a decision analysis from scratch, the initial display shows only a root decision node and one simple choice branch, as in Figure 1. A pop-up menu can be used to select among the commands that display when a node is clicked. On the root node, the options are "Edit Label" and "Create Choice." Create Choice is used to add a new subtree element under the root node, another possible decision choice. Other nodes have a "Delete" option, but the root node cannot be deleted, only renamed.

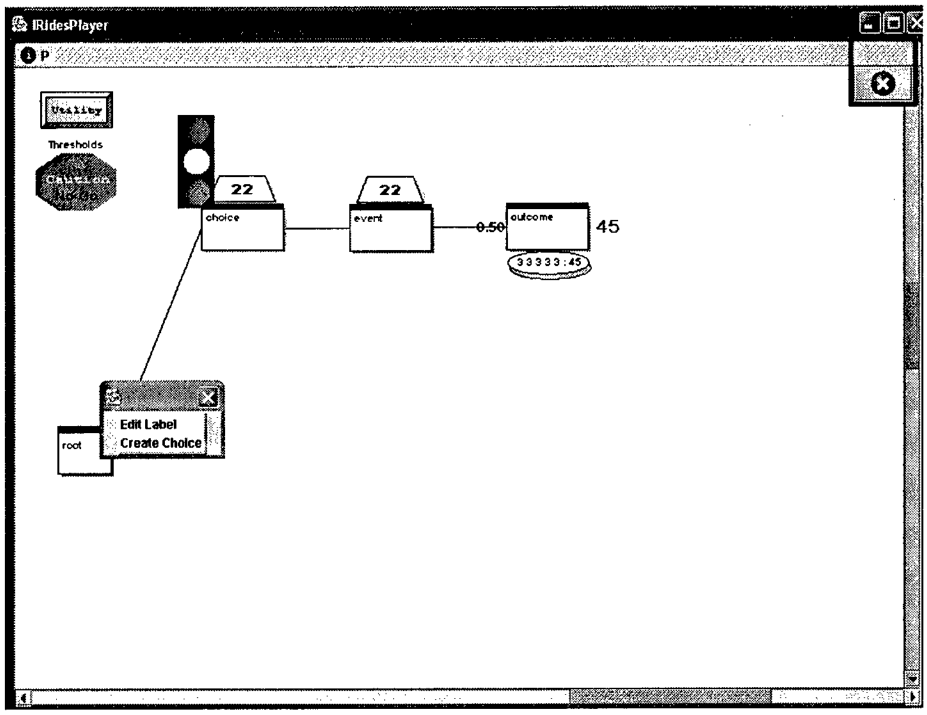


Figure 1. "Empty" DAT model.

Authors relabel the nodes to reflect the choices in the context being analyzed and the possible outcomes of decisions. They can also create new nodes, including additional choices, events, and outcomes. At the development point shown in Figure 2, the original nodes have been relabeled and the author has created two possible outcomes for the first evaluation: a good result and a poor one.

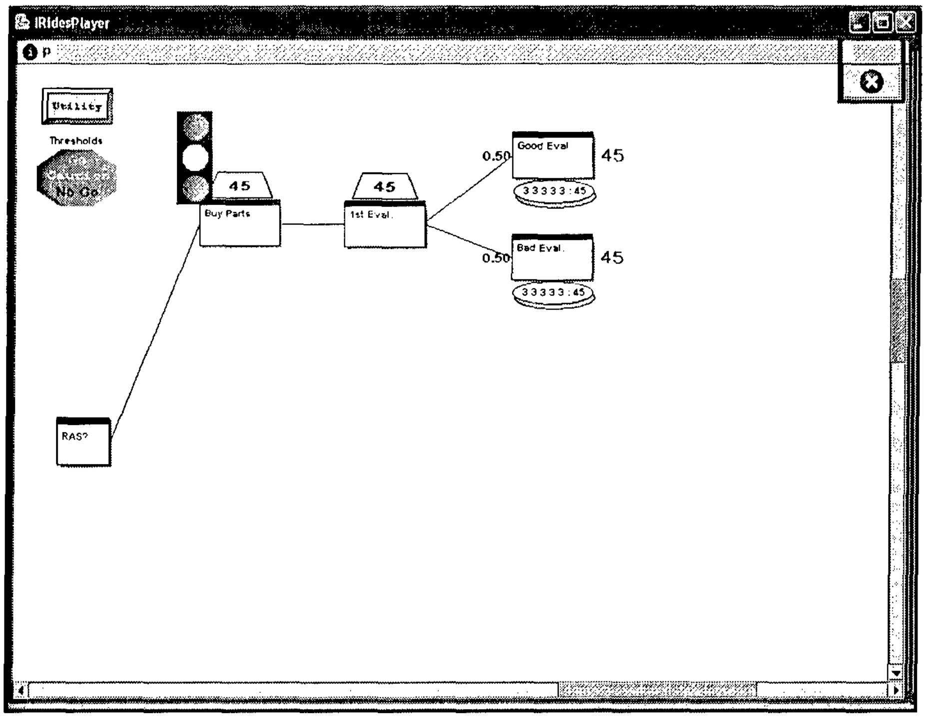


Figure 2. Renaming nodes, creating a new outcome.

For a given decision domain, outcomes have appropriate *attributes*. Authors can enter the names of the attributes that apply to the decision that is being analyzed. Clicking on the *Utility* button opens the Attributes Definition interface (Figure 3). In the original, empty DAT document, there are five utility attributes named "a" through "e." Initially, each defaults to an intermediate *factor* of 3. These factors are the weights by which actual attribute values of particular outcomes are multiplied to compute the total value or utility of each outcome.

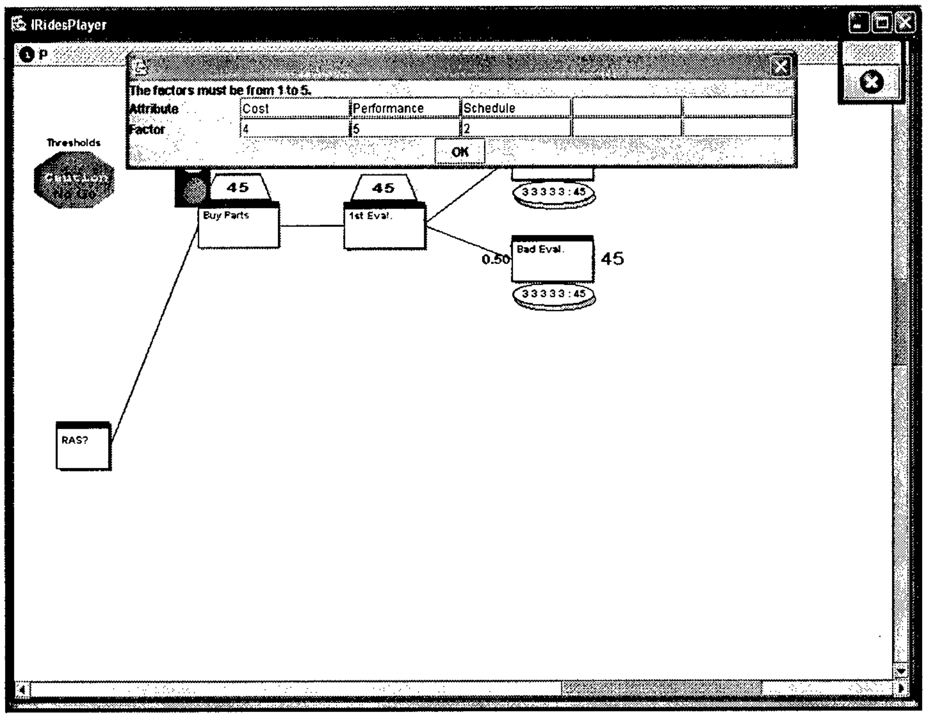


Figure 3. Defining attributes.

When the Attributes Definition interface is closed, the number of attribute values displayed below each outcome node is updated, if any attributes have been deleted or if new ones have been added. Because the original attribute names "d" and "e" were deleted, in Figure 4 there are only three attribute value numbers below each outcome, although there were five in the earlier figures.

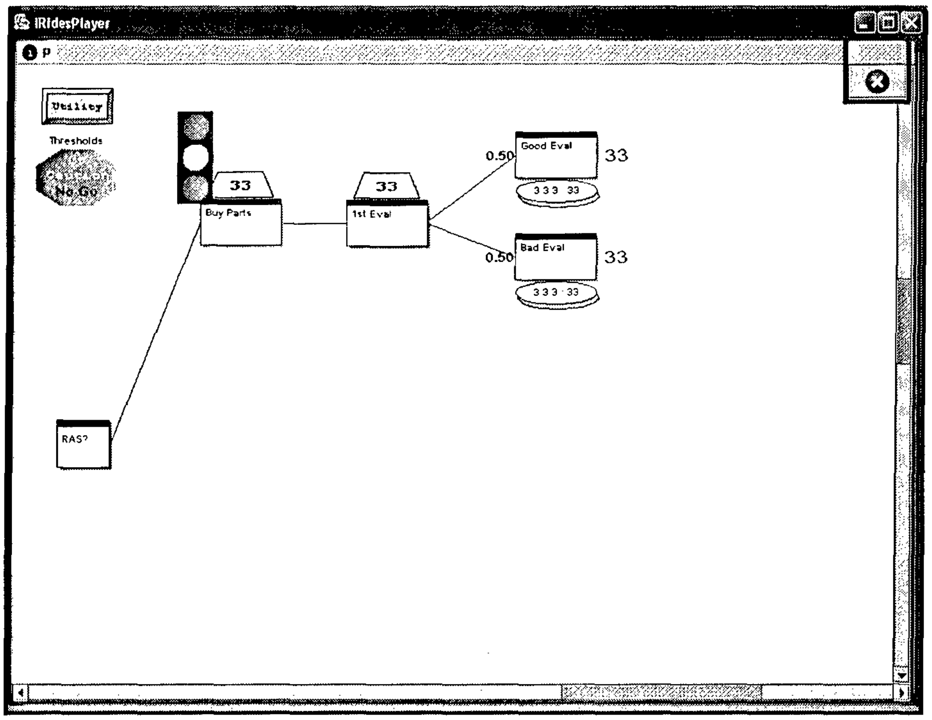


Figure 4. Updated number of attributes in outcomes.

Clicking on the attribute values of an outcome node opens the Attribute Settings dialog. For each outcome, the user can specify how good or bad the result will be in terms of each attribute. In the case shown here, the Cost result will be neutral (3) if the parts inventory is purchased and a good evaluation results. The Performance will be excellent (5), and the Schedule will also be excellent, because roughly half of the planned production run will be completed. As these values are selected in the dialog, the numbers change in the outcome's ellipse in the main screen, and expected values are also automatically recomputed.

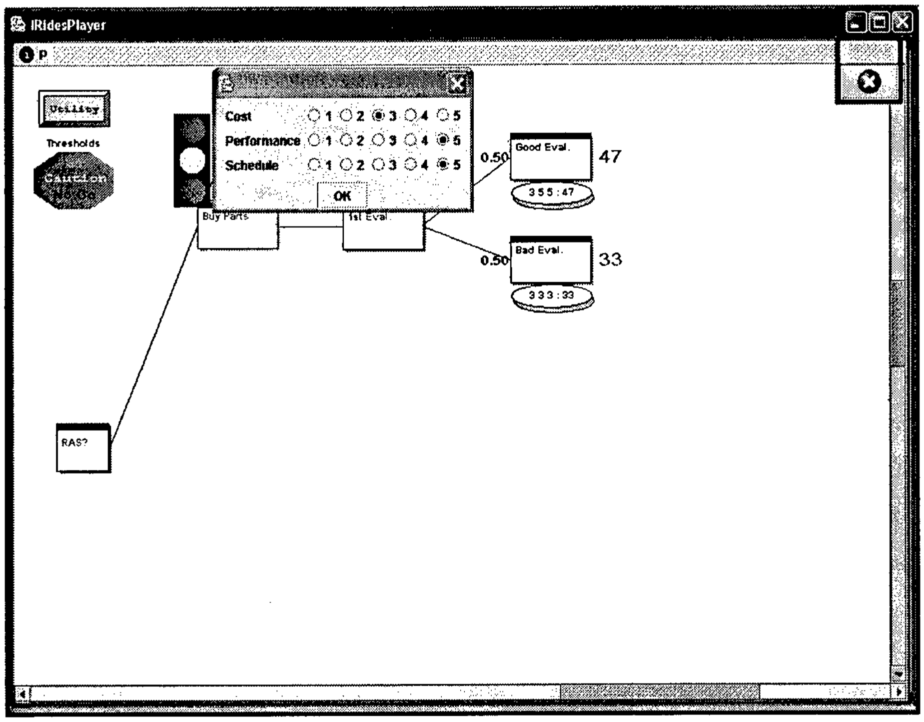


Figure 5. Specifying the attribute values of an outcome.

Not every outcome of a post-choice event is equally probable. The expected value of a choice is dependent not only on the utility of resulting incomes, but also on the probability those outcomes will occur. Estimated probabilities are shown as numbers just to the left of outcome nodes. When new outcomes are first created, they are equally likely. (Note the 0.50 values to the left of the outcome nodes in Figures 2–5.)

Clicking on a probability opens a Probability slider. In Figure 6, the author has decided that there is a 4% chance of a poor first evaluation after making the "Buy Parts" decision. As the slider is dragged to a new value, the corresponding alternative outcome's probability is automatically altered so that the numbers sum to one. (If an event has three or more possible outcomes, the probability of each outcome must be set manually.)

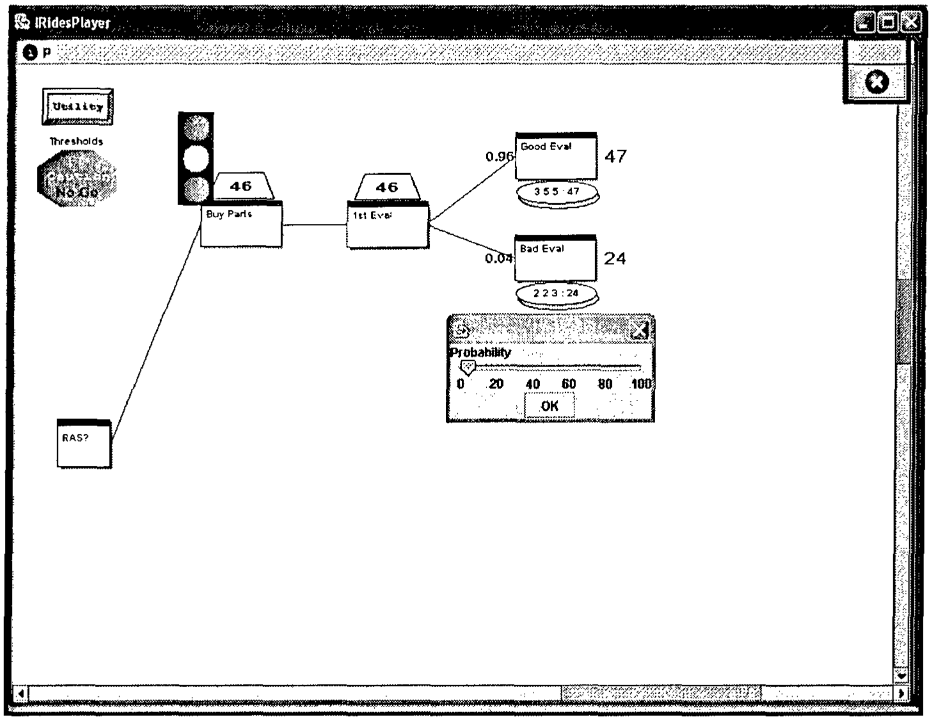


Figure 6. Assigning probabilities to outcomes.

By continuing to add choices and outcomes, editing the utility values, and specifying estimated probabilities, a user can develop a rich representation of many aspects of a problem. In Figure 7, the values selected by the user do not result in large differences in the expected values of the choices analyzed. The traffic light signal shown to the left of each choice node reflects the "go–caution–no-go" presentation approach advocated in the EDO Basic Course. In each case here, the three choices are marked here with the yellow "Caution" symbol at the center of each stoplight.

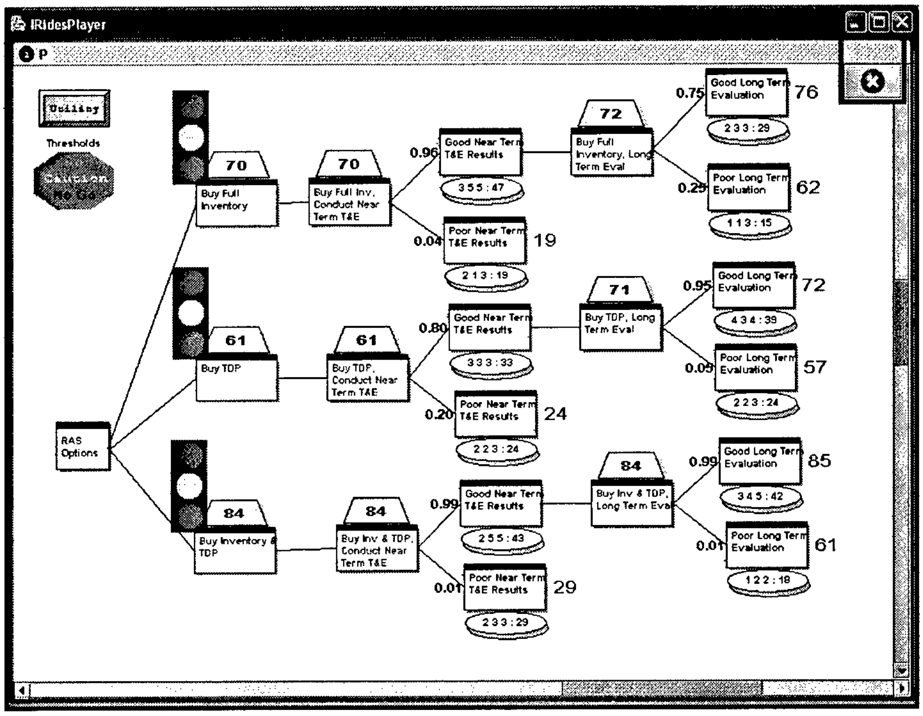


Figure 7. A nearly complete decision analysis.

It is possible to manipulate the thresholds of the signals using a pair of sliders. Clicking on the button labeled "Go–Caution–No-Go" reveals the slider interface, as shown in Figure 8. Here, the user has slightly lowered the threshold for "Go" by dragging the green/yellow (top) slider down a bit. The third choice is now marked with a green light as the one to be preferred. Depending on how the students set the thresholds, all, some, or none of the possible courses of action they propose may produce "acceptable" outcomes.

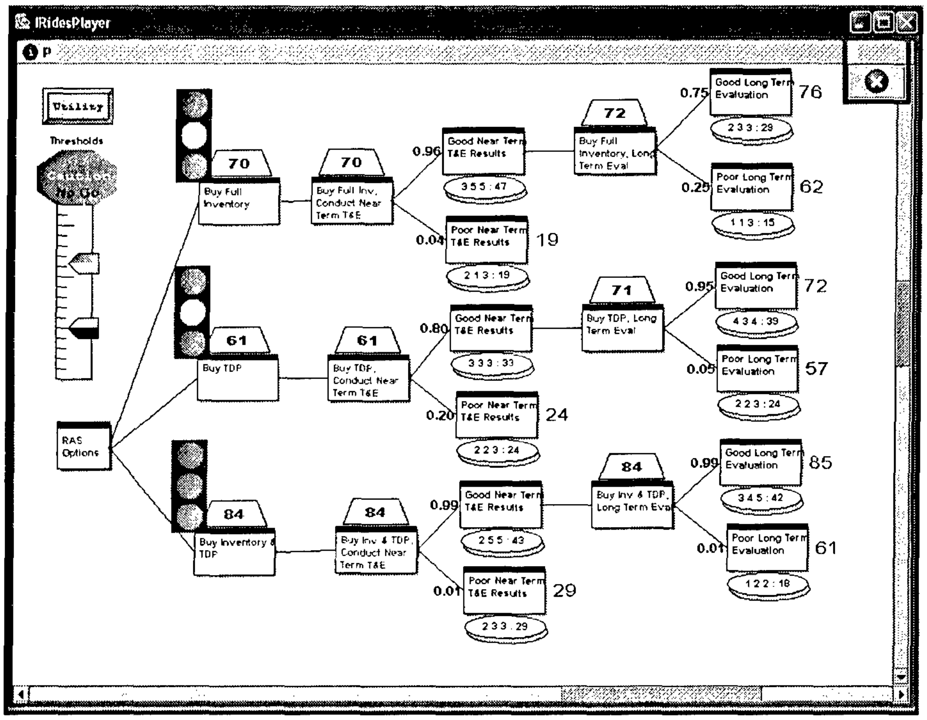


Figure 8. Adjusting acceptance and rejection thresholds.

The RAS partial analysis. When students begin working on the Refueling At Sea problem, they open a DAT model that includes three obvious choices (the three shown in Figure 7), but with all outcomes having equal probability and equal utility. They modify these estimates to create more nearly complete analyses. They can also delete choices that they believe are not worthy of consideration, and they can devise and insert new procurement options of their own design.

Recording DAT usage data. There are seven types of events generated when students use the tool (create a new course of action, delete a course of action, label a course of action, weight the overall importance of utility attributes relative to one another, set the probability that each outcome will occur, set the utility attribute(s) of each outcome, and set the threshold values). In addition, the system will generate a "stoplight" event whenever a student action causes the expected value of a possible outcome to cross the student-determined threshold (e.g., when the value of the decision moves from acceptable to marginal or from unacceptable to acceptable).

When using the Decision Analysis Tool, each action performed by a student is recorded in an electronic "clickstream" file. Each file entry includes a student identifier, the date and time of the event, the action that generated an event and the target of the event, and the value assigned to the target of the event. For example: "EDO 33, Monday Feb 09 2004 14:45, label option 1, 'Buy Full Inventory, $20 M.'"

Because students may generate an unlimited number of procurement actions and students can evaluate an unlimited number of future decisions made about each action, we cannot identify the specific objects students will create or how they will manipulate those objects before each use of the tool. Consequently, it is difficult to evaluate a student solution by comparing it to a single correct solution or to compare one student solution to another student solution. Furthermore, even if we defined the largest solution space developed by any student to date as the basis of comparison, the number of degrees of freedom combined with the small number of clicks on less often chosen targets would make meaningful parametric or non-parametric (such as Artificial Neural Network) analysis difficult. Finally, the instructors at the EDO school have indicated that they do not require such a detailed analysis.

However, as discussed above, the number of event types generated by a student remains fixed at seven regardless of the number of objects a student creates. This allows us to reclassify the more granular data into more general "click type" categories. Not only does such a reclassification have the benefit of providing a more useful level of data analysis to EDO instructors, reclassifying events at this more macroscopic level has the additional advantage of reducing the degrees of freedom to a level necessary to allow us to apply appropriate data analysis methods.

Analysis of Decision Analysis Tool Data

After reclassification, we perform a frequency analysis of the clickstream data. This analysis allows us to determine how much time a student spent using the tool, the total number of interactions (clicks) a student had with the tool, and how those interactions were distributed among the seven event types. This later analysis is particularly important for identifying students who concentrate on (or ignore) specific aspects of a decision. For example, students who do not adjust the probability values of an outcome may not be considering the impact of the uncertainty of forecasts in their decisions, whereas frequent changes in utility or probability values can suggest students are unclear about these concepts, are performing a sensitivity analysis, or are trying to justify a specific outcome. Frequency analysis also allows us a rough comparison of how various students navigated the problem space and how the observed interaction with the system relates to some expected model of interaction. Finally, frequency analysis provides us a guide as to which event types occur so infrequently they can reasonably be

ignored in subsequent sequential analysis. For example, we suspected that creating new decision paths, while an important characteristic of some solution strategies, would occur so infrequently compared to other click events that their presence (or absence) alone, rather than how often they occurred, would make a solution strategy distinctive.

Frequency analysis, however, has certain limitations. For one, frequency analysis is only able to tell us that a decision maker manipulated certain parameters in the problem space. It tells us nothing about the order in which those events occurred. Consequently, frequency analysis allows us to say little about the process that an EDO student took to reach his or her final decision. It is conceivable that the frequency of clicks could be identical for two different student groups, but that the sequence of clicks is entirely different. For example, one process might involve adjusting the probabilities of an event occurring for all events in the solution space and then determining the utility of each event. Another, very different, strategy might be to sequentially adjust the utility and probability of each event in the problem space. While frequency analysis would suggest the two processes were identical, a sequential analysis would immediately identify the differences because it would highlight the patterns identified by the processes. This type of sequential modeling should also allow us to associate particular processes (or parts of a process) with outcomes and, perhaps, to identify shortcomings in the final solution or the process to reach such a solution. As a demonstration of the concept, we applied sequential analysis to the EDO data. Although we consider only pairs of clicks in the present analysis, we could apply this same methodology to larger data sets in order to determine the significance of click triplets, quadruplets, etc.

Although we expected probability and utility clicks would prove to be the most distinguishing characteristics of student solution strategies because of their importance to the underlying theories and their predominance in the solution area, we wanted to use our initial studies in a more exploratory fashion and not restrict our analysis to these variables alone. In exploratory analyses, Bakeman and Gottman (1997) argue that traditional lag-sequential analysis invites type 1 error, especially when the number of event types is large. For example, with 10 event codes, we would expect 5 out of 100 precedent-antecedent click-pairs to appear statistically significant at the .05 level of significance when no significant relationship actually exists. To minimize this error, Pearson's Chi-square (χ^2) is calculated for the results of each student group to determine if prior and antecedent clicks are significantly

associated (or are independent of one another) at $\alpha = .05$. In addition, we limit our investigation to those events that occur more than five times in the data, again as suggested by Bakeman and Gottman (p. 145). While the Likelihood-ratio χ^2 (or G^2) statistic is often preferred in sequential analysis, Pearson's Chi-square is favored when sample sizes are small as in this study (Bakeman & Quera, 1995). Next, we compute z-scores (adjusted residuals) for each two-event sequence to identify those transitional probabilities that are significant at the .05 level. While raw residuals are used in the Pearson Chi-square statistic, their reliance on the number of times a certain pair occurs adversely impacts their usefulness in comparing problem-solving performances that do not have the same total number of clicks. Consequently, we relied on adjusted residuals merely to check for conditions that might cast doubt on the normality of the distribution of residuals. Unless otherwise noted, we limit our discussions to those cells where we have the data to assume residuals are normally distributed. We also compute conditional transitional probabilities for each click-pair in order to determine the probability that a second click event follows a given prior event. It is important to note that these conditional transitional probabilities do not merely represent the probability that a certain click-pair occurred out of all click-pairs. As Bakeman and Gottman argue, in sequential analysis we are interested in what click followed some previously defined click. Therefore, we must account for the frequency of the prior events in the data set in order to "correct" for priors that occur with differing frequencies (also known as base rates). Finally, given the sensitivity of z-scores and Chi-square results to sample size, we compute a Yule's Q statistic to compare the magnitude of an effect for each click-pair across group samples. The Yule's Q statistic is based on the odds ratio (how much more likely is it that an event will occur than not occur). Consequently, it is unaffected by sample size. Yule's Q is +1 when a specific type of click always predicts the click that will follow or when one can conclude that in the absence of a certain type of click, a specific type of click will not follow. Yule's Q is -1 in the exact opposite situation, that is when a specific type of click predicts that the click to follow will not be of a specific type or that in the absence of a certain type of click a specific type of click will follow. When Yule's Q is zero, there is no relation between precedent and antecedent clicks. Here again, because we are most interested in examining click-pairs that are likely to co-occur in solution strategies, we focus on large, positive values of Yule's Q ($\geq .75$). Additional analysis could be performed by focusing on values of Yule's Q that are nearly zero, or are large and negative.

Student Survey

In order to evaluate the accuracy of some of our inferences about the student beliefs which prompted certain behaviors and to evaluate the impact of the tool on student learning, we administered a 15-item survey to each student. The survey asked students for their perceptions of their level of knowledge about the domain, impact of the Decision Analysis Tool, and general impressions about the DAT and Knowledge Mapping Tool. The survey is given in Appendix F. We then correlated these results with the results obtained from the analyses described in previous sections.

Results

Human Performance Knowledge Mapping Tool

One finding from the pilot test indicated that there was confusion about the meaning of the links. That is, the representation is inherently procedural and the knowledge is in the flow of events. Requiring students to specify reasons for transitions appeared to severely interfere with constructing maps. The second finding is that students rarely used the type-in option. Third, there was consensus among the instructors and CRESST researchers that the level of specificity of the map content was too narrow. That is, the content of the map covered only a small segment of the curriculum and would be of little use.

A paired *t* test was conducted to check for pretest-posttest differences. While there was a difference in knowledge map scores that showed students' map scores trend higher, this difference was not statistically significant at the .05 level ($p = .11$). As shown in Figure 9, the distribution of knowledge map scores shows that the scoring procedure yielded a range of scores, and the scores appear to increase from the pretest to the posttest. This result is consistent with prior work but its measurement quality remains unknown—there was no referent available (e.g., course grades or other outcome measure) to compare knowledge map scores against.

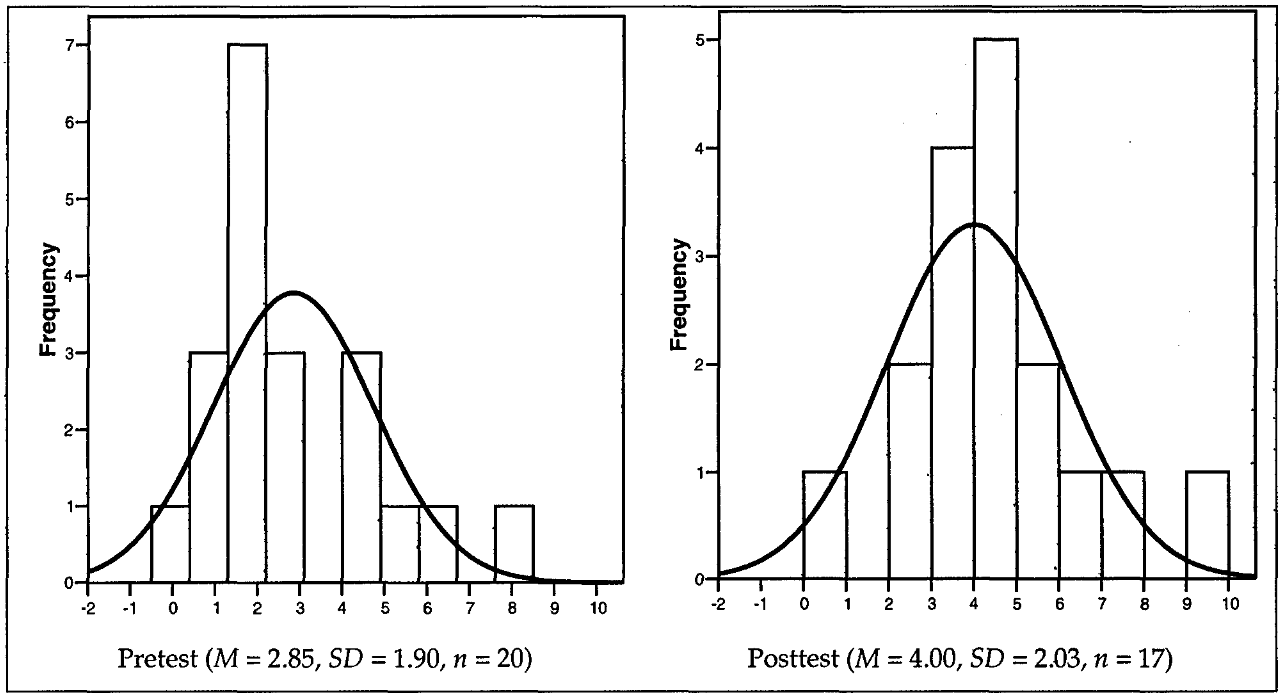


Figure 9. Distribution of knowledge map scores.

Some insight into whether knowledge mapping is potentially useful as a tool for measuring understanding of process can be seen in the knowledge maps for different students. Figure 10 shows an example of the lowest scoring map (map score = 0). Note the map depicts the problem-solving process in a nonsensical way. In particular, the step "define the environment" is treated the same as other steps; however, defining the environment is a key driver in any analysis of risk—the environment defines the operating conditions and governs the set of constraints that will be most applicable to the current problem. In contrast, the highest scoring map shown in Figure 11 shows a clear and logical sequence of steps. Finally, Figure 12 shows the pretest and posttest maps of the student with the highest gains. The pretest map omits an important step (analysis of alternatives) and depicts legal constraints driving the process. The posttest map is a much more refined map.

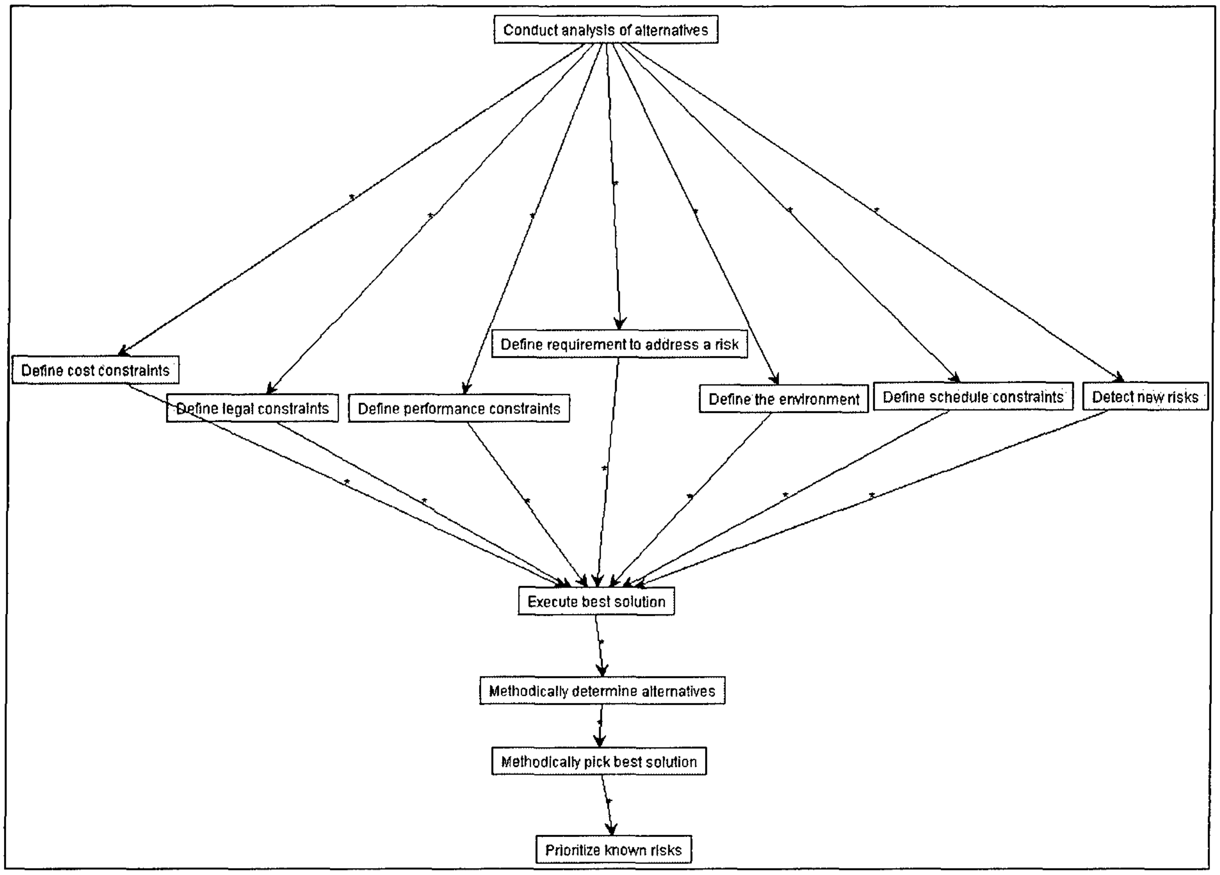


Figure 10. Example of a low-scoring knowledge map (pretest, score of 0, [EDO111]).

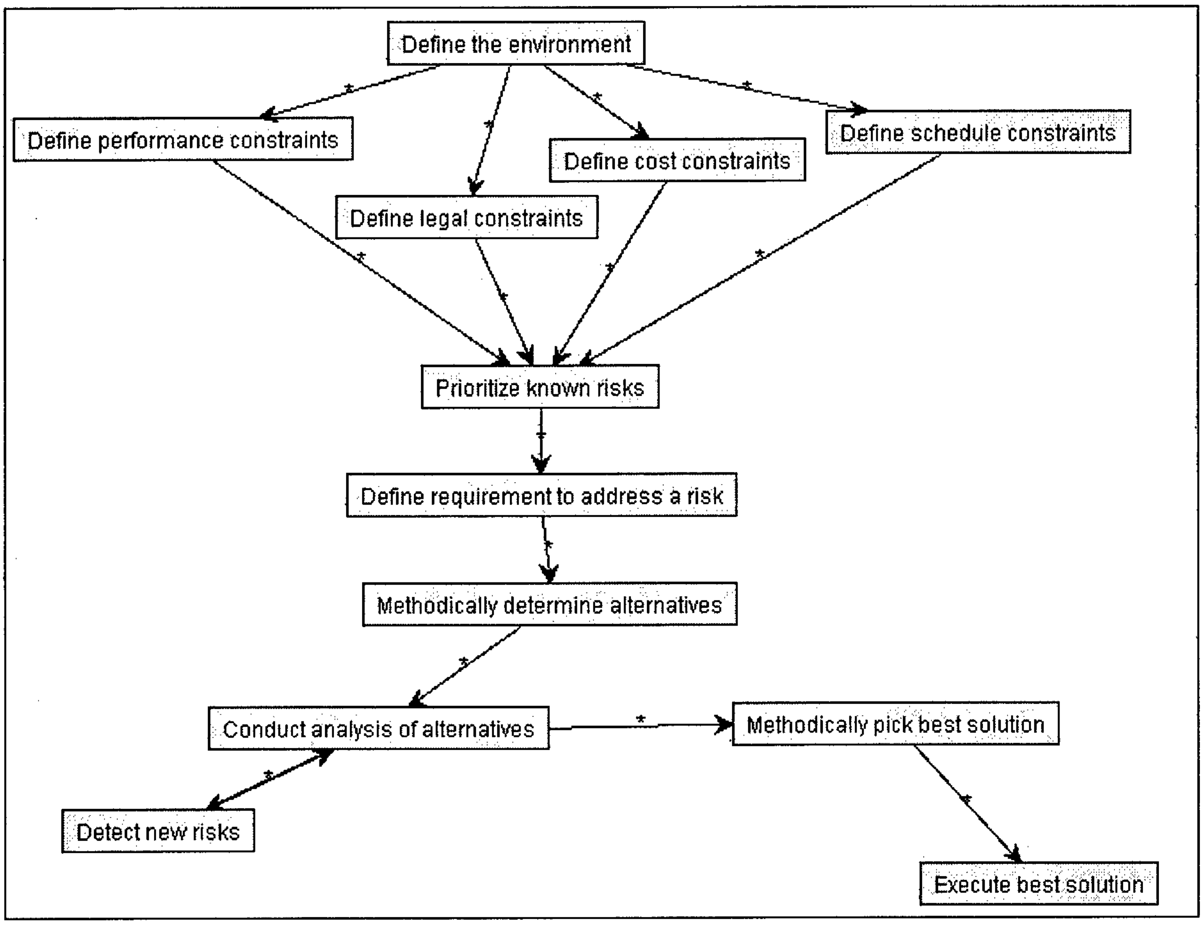


Figure 11. Knowledge map of highest scoring knowledge map (posttest score of 9, [EDO108]).

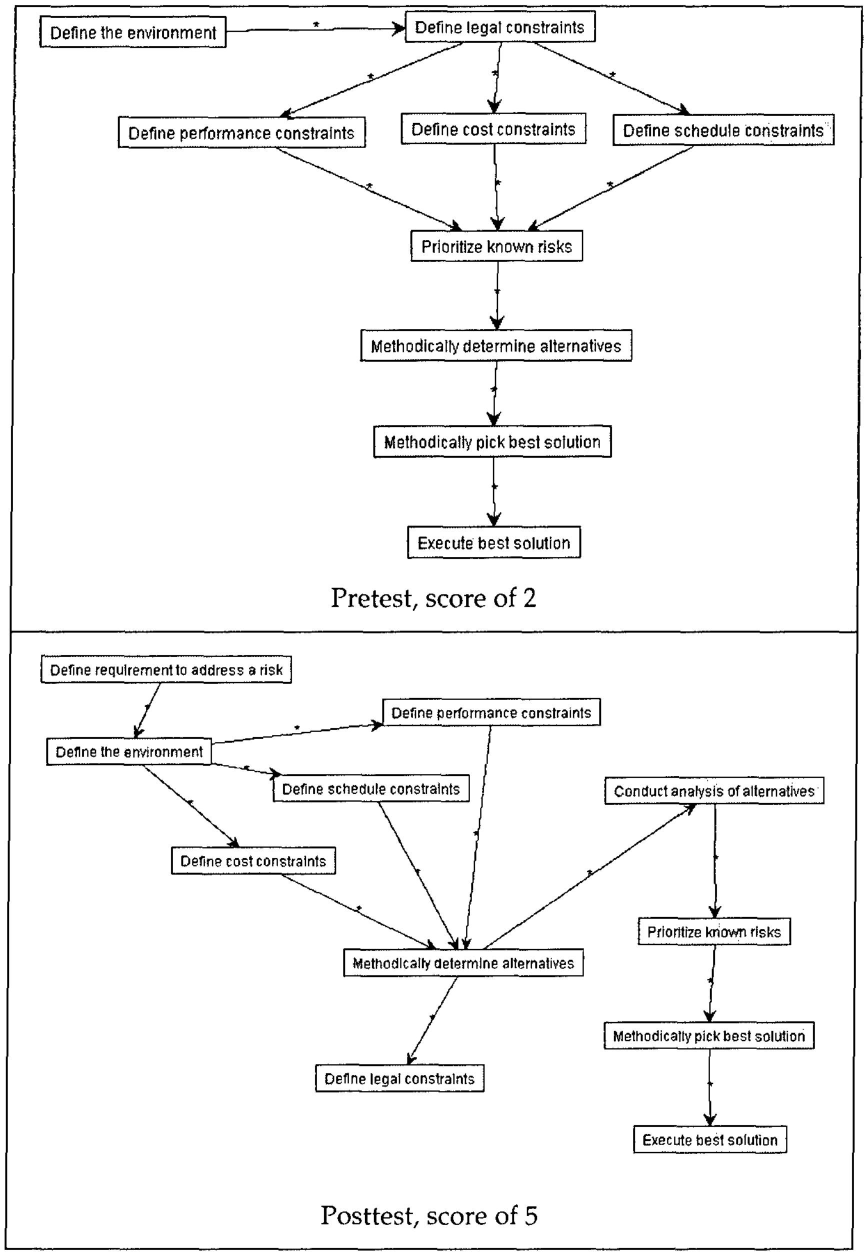


Figure 12. Knowledge maps of student demonstrating a large gain in map scores from pretest to posttest (EDO115).

Decision Analysis Tool

As described above, the final evaluation of students in the EDO course requires each group of approximately six students to present their proposal for the U.S. Navy to acquire a new frigate. This presentation represents approximately 10% of each student's final course evaluation. The presentation is designed by the course instructors to be a Milestone B review and requires each group to consider the cost, performance, and schedule of each major subsystem in the program. One of the

subsystems that the students must address in their presentation is the RAS system, how they plan to respond to programmatic changes and uncertainty, and how they will generate and evaluate alternatives in order to accommodate these changes. While the student groups have an extended period of time to develop their acquisition proposals, the RAS problem is given to them only 48 hours before their final presentation. In addition, other course requirements take up a large portion of the students' days, which leave them a limited time to develop their presentations. Consequently, most groups develop their proposals and their resolution of the RAS problem in parts, with small subteams working largely independently on their respective part of the presentation. As a result, only one or two team members generally used the Decision Analysis Tool when evaluating the group's solution to the RAS problem. The reader should keep this caveat in mind in the analysis that follows.

We collected the data for this analysis from two classes of EDO students. Groups 1–3 represent student groups who used the tool in 2003, and Groups 4–6 were students in a 2004 class.

Decision Analysis Tool Frequency Data—Time and Rapidity of Interaction

The data in Table 1 describe the frequency of each type of event generated by students using the Decision Analysis Tool, as well as the length of time students used the tool in preparing their presentation. On average, the groups generated events in the tool every 2 or 3 minutes; however, the total number of interactions with the tool and the total amount of time each group used the tool varied quite dramatically. Groups that spent less time with the tool interacted with the system more frequently than groups that spent more time (3.3 clicks per minute on average as compared with 2.5 clicks per minute) and the difference was significantly different between the three groups that used the tool in 2003 and the three groups that used the tool in 2004 (t = -3.3, p = .03). The frequency of each click type also varied between the student groups.

Table 1

Frequency Distribution of Events by Group

	Group 1		Group 2		Group 3		Group 4		Group 5		Group 6	
Event type	Freq.	% of total	Freq.	% of total	Freq.	% of total	Freq.	% of total	Freq.	% of total	Freq.	% of total
Utility	157	57%	84	51%	33	41%	89	50%	52	66%	81	48%
Stoplight	52	19%	39	23%	19	23%	30	17%	6	8%	54	32%
Probability	34	12%	13	8%	9	11%	18	10%	12	15%	16	9%
Thresholds	18	7%	11	7%	12	15%	14	8%	0	0%	8	5%
Labeling	2	1%	8	5%	4	5%	16	9%	5	6%	3	2%
Weights	3	1%	7	4%	0	0%	0	0%	3	4%	3	2%
Total clicks§	275	100%	166	100%	81	100%	177	100%	79	100%	170	100%
Total time	96 mins.		75 mins.		32 mins.		55 mins.[ξ]		25 mins.		47 mins.[ξ]	
Clicks/min.	2.9		2.2		2.5		3.2		3.2		3.6	

[ξ]Both Group 4 and Group 6 had periods of inactivity (20 minutes and 87 minutes, respectively) in their interaction with the DAT. These periods have not been included in the total time value for either group.

§Total clicks includes create and delete event clicks, but these are not disaggregated in the above table.

Decision Analysis Tool Frequency Data—Utility, Stoplight, and Probability

In all classes, three or four of the seven possible events account for the vast majority of activity in the decision-making process. Given that the stoplights indicate a "go–caution–no-go" decision for each branch of the decision tree, and that utility, probability, and thresholds are the three main determinants of the stoplight colors, this is not surprising.

All the same, given that changes to utility, probability, and thresholds each impact a determination of which decision path is optimal, the students in all but two groups focused on utility significantly more often than would be expected if their actions were distributed in proportion to the click areas in the problem space (Group 1: $\chi^2 = 14.45, p < .001$; Group 2: $\chi^2 = 19.2, p < .001$; Group 3: $\chi^2 = 52.15, p < .001$; Group 4: $\chi^2 = 20.4, p < .001$; Group 5: $\chi^2 = 4.38, p = .11$; Group 6: $\chi^2 = 7.57, p = .02$). Group 3 focused less on utility than would be expected, and Group 5's click pattern was not significantly different than would be expected, although the lack of statistical significance could be due to this group's relatively low number of total clicks. For

every group, clicks that involved adjustments to the utility of a decision were the primary focus of the group's attention, and only one group (Group 3) adjusted utility values significantly less than half the time. Two groups (Group 1 and Group 5) adjusted these values significantly more than half the time. A full two thirds of the interactions Group 5 had with the system involved adjustments to the utility parameters of their decisions.

In all but one group (Group 5), the number of stoplight events accounts for the second most commonly occurring events in the students' performances. In one case (Group 6) the number of stoplight events represents almost one third of all the events generated and differs greatly from the number of these events in every other group.

The third most frequently occurring events in the data set for all but two groups (Groups 3 and 5) are events that involve adjusting the probability estimate of a future event. As explained above, this is essentially the student's forecast that some event will occur. In general between 8% and 12% of the groups' clicks involved adjustments to the probability estimates. The exception to what we generally observed is evident in Group 5's performance, where 15% of their click interactions involved setting probability parameters. As discussed below, Group 5's strategy is very different from the other strategies in at least one major way.

Groups 3 and 5 are also noteworthy in the percentage of their interactions that involved setting threshold values. Thresholds determine where stoplights will indicate a "go–caution–no-go" signal. It is unusual for groups to merely accept the preset values when making a decision, rather (as explained below) they generally either set these in response to or in order to trigger stoplight events. Whereas Group 3's extraordinarily high percentage of changes to thresholds suggests a focused examination of the actual clickstream data might be in order, that Group 5 never adjusted these thresholds only invites speculation as to why no changes were made.

Decision Analysis Tool—Lag Sequential Analysis

We used the lag sequential analysis method described by Bakeman and Gottman (1997) to investigate dyads of clickstream sequences. As noted above, this analysis could be easily extended to clusters of three or more clicks if required.

In each case below, Pearson's Chi-square indicates that the precedent-antecedent clicks are associated ($p \leq .001$), so we calculated z statistics for each cell of

the precedent-antecedent click table at the .05 level of significance. Because the total number of clicks in most performances is small (< 200), we chose to discount the significance of cells where we could not assume that residuals were normally distributed. Finally, since we focus on the clicks that followed (versus the clicks that *did not* follow) a precedent click, we highlighted cells that were not only significant but that had a relatively large Yule's Q (≥ .75).

Group 1

In deciding which RAS solution was "best," the students in Group 1 adjusted all the utility values in the problem space one after another, and then did the same with all the probability values. Only an occasional stoplight event interrupted their very sequential click pattern. The majority of stoplight events in this group's performance, however, did not result from adjustments to utility or probability events.

Table 2

Conditional Transitional Probabilities (Group 1)

Given \ Target	Probability	Stop	Utility	Threshold
Probability	.91**	.08	.00**	.00
Stop	.03*	.56**	.10**	.31
Utility	.01**	.04**	.96**	.00
Threshold	.00	.77	.11	.11

*$p < .05$ and normally distributed residuals can be assumed.

**$p < .01$ and normally distributed residuals can be assumed.

Table 3

Yule's Q Statistics (Group 1)

Given \ Target	Probability	Stop	Utility	Threshold
Probability	+1.00	-.48	-1.00	-1.00
Stop	-.64	.82	-.92	.96
Utility	-.97	-.90	.99	-1.00
Threshold	-1.00	.90	-.87	.28

The .96 Yule's Q in the Stop-Threshold pair results from the fact that, with only two exceptions, this group of students made threshold adjustments only after "prompting" by stop events. Similarly, the students almost always generated a stoplight event when adjusting the thresholds. While the number of occurrences of these pairs does not allow us to conclude either relationship is statistically significant, the data do show that stoplight changes resulted almost entirely from changes made to the threshold sliders and not from changes in the expected value of various procurement alternatives.

Group 2

In many ways, Group 2's solution to the RAS problem was similar to the strategy used by Group 1. Both groups focused on making changes to the utility values associated with each procurement alternative.

Table 4

Conditional Transitional Probabilities (Group 2)

Given \ Target	Probability	Stop	Utility	Threshold
Probability	.23	.31	.46	.00
Stop	.08	.56**	.11**	.25
Utility	.09	.09**	.83**	.00
Threshold	.00	.50	.30	.20

**$p < .01$ and normally distributed residuals can be assumed.

Table 5

Yule's Q Statistics (Group 2)

Given \ Target	Probability	Stop	Utility	Threshold
Probability	.56	.14	-.25	-1.00
Stop	-.07	.75	-.91	.89
Utility	-.10	-.82	.89	-1.00
Threshold	-1.00	.53	-.55	.54

As suggested by the .89 value of Yule's Q in the fourth column of Table 5, almost every utility event was followed by another utility event. Also, like Group 1, this group almost always made threshold adjustments after a stoplight event occurred. As before, the small number of these Stoplight-Threshold pairs does not allow us to make a claim of statistical significance. Two characteristics, however, make Group 2's approach to procuring a RAS capability notably different from Group 1's strategy. The most notable difference between the strategies of the two groups is this group's integration of probability events within the milieu of all the other events in the group's clickstream. In addition, unlike Group 1, this group did not routinely generate stoplight events when changing threshold values.

Group 3

Like the previous two groups, Group 3's strategy focused heavily on adjusting the utility values of the procurement options within the solution space. Similarly, this group reacted to changing stoplight values by changing the threshold levels for the "go–caution–no-go" on the stoplight rather than going back to change utility or probability values.

Table 6

Conditional Transitional Probabilities (Group 3)

Given \ Target	Probability	Stop	Utility	Threshold
Probability	.22	.22	.56	.00
Stop	.06	.44	.06	.44
Utility	.16	.06**	.74**	.03**
Threshold	.00	.55	.27	.18

**$p < .01$ and normally distributed residuals can be assumed.

Table 7

Yule's Q Statistics (Group 3)

Given \ Target	Probability	Stop	Utility	Threshold
Probability	.44	-.12	-.21	-1.00
Stop	-.46	.53	-.93	.86
Utility	.38	-.83	.81	-.83
Threshold	-1.00	.64	-.45	.09

In this performance, only the 74% probability that a utility click will be followed by another utility click is significant with a large Yule's Q, but this probability is much lower than for either Group 1 or Group 2. The data suggest this group integrated their utility and probability clicks more than most of the other groups. In fact, the magnitude of the Yule's Q measure for the Utility-Probability transition is positive only for this group. We note that because Utility-Probability clicks are less common in the data set, we would not expect to see this pair a high percentage of the time, so the conditional transitional probability for the pair is still smaller than one might expect.

Group 4

In a manner similar to previous groups, the strategy Group 4 used to solve the RAS problem relied largely on sequentially manipulating utility variables in the problem space.

Table 8

Conditional Transitional Probabilities (Group 4)

Given \ Target	Probability	Stop	Utility	Threshold
Probability	.56	.28	.17	.00
Stop	.13	.37**	.07**	.40
Utility	.05**	.05**	.90**	.01
Threshold	.00	.71	.21	.07

**p < .01 and normally distributed residuals can be assumed.

Table 9

Yule's Q Statistics (Group 4)

Given \ Target	Probability	Stop	Utility	Threshold
Probability	.91	.29	-.77	-1.00
Stop	.12	.55	-.93	.95
Utility	-.67	-.84	.96	-.90
Threshold	-1.00	.88	-.68	-.12

As was the case with Groups 1 and 2, the .95 Yule's Q in the Stop-Threshold pair occurs because, with only two exceptions, this group of students made threshold adjustments only after "prompting" by stop events. Similarly, the students almost always generated a stoplight event when adjusting the thresholds. Finally, the high Yule's Q for the Probability-Probability event pairs suggests that this pair occurs often for this group, and (along with the high Yule's Q for Utility-Utility pairs) suggests that the group was very sequential in their approach to solving the RAS procurement. The lower than anticipated conditional transitional probabilities for Probability-Probability pairs is due to the large number of stoplight events generated as the group changed probability values. These stoplight events result from the fact that when this group made changes to probability values, the changes were often large enough in magnitude to generate stoplight changes, which lowers the Probability-Probability conditional transitional probabilities. Nevertheless, Yule's Q was able to highlight the Probability-Probability pair as important because the student generally clicked on a probability slider only after they had made

adjustments to some probability value immediately prior. As with other groups, however, the number of occurrences of both the Probability-Probability and Stop-Threshold pairs does not allow us to conclude either relationship is statistically significant.

Group 5

As suggested in the review of the DAT frequency data above, Group 5 approached the RAS problem with a solution strategy that was similar to Group 3 in many ways.

Table 10

Conditional Transitional Probabilities (Group 5)

Given \ Target	Probability	Stop	Utility	Threshold
Probability	.42	.08	.50	.00
Stop	.50	.16	.33	.00
Utility	.06**	.08	.82**	.00
Threshold	.00	.00	.00	.00

**$p < .01$ and normally distributed residuals can be assumed.

Table 11

Yule's Q Statistics (Group 5)

Given \ Target	Probability	Stop	Utility	Threshold
Probability	.69	.00	-.47	Undef
Stop	.73	.42	-.68	Undef
Utility	-.85	-.11	.77	Undef
Threshold	Undef	Undef	Undef	Undef

Although the Yule's Q and conditional probability values are still relatively high and are significant, this Yule's Q has the lowest values for the Utility-Utility click-pair of any of the six groups in this study. Practically, this suggests that this group had one of the most "integrated" approaches to the RAS solution in that they

incorporated utility and probability when considering the efficacy of each option individually. However, that this group never manipulated threshold values is somewhat perplexing and suggests that they may have never considered what defined a "go–caution–no-go" decision for themselves.

Group 6

The strategy of Group 6 closely parallels that of Group 1. As with Group 1, Group 6 used a strategy that focused on certain types of events in a sequential manner. Unlike Group 1, however, Group 6 integrated small chains of probability events into their changes in utility values.

Table 12

Conditional Transitional Probabilities (Group 6)

Given \ Target	Probability	Stop	Utility	Threshold
Probability	.50	.06	.38	.00
Stop	.04*	.73**	.08**	.15
Utility	.08	.08**	.85**	.00
Threshold	.00	.71*	.29	.00

*$p < .05$ and normally distributed residuals can be assumed.

**$p < .01$ and normally distributed residuals can be assumed.

Table 13

Yule's Q Statistics (Group 6)

Given \ Target	Probability	Stop	Utility	Threshold
Probability	.88	-.78	-.30	-1.00
Stop	-.60	.91	-.94	+1.00
Utility	-.29	-.89	.93	-1.00
Threshold	-1.00	.70	-.47	-1.00

While this suggests some similarity to Group 2's strategy, Group 6 often performed a series of probability adjustment clicks rather than a single adjustment

in the midst of a string of adjustments in utility variables. Like a number of groups discussed earlier, Group 6 only adjusted the threshold values in their solution space after stoplight events prompted them to do so. This group also generated significantly more stoplight events than any other group. Like Group 1, the students in this group often generated a stoplight event when adjusting the thresholds, but there is a significant difference between Group 6 and all other groups with regard to threshold adjustments. The data here, along with our observation that many of the threshold events occur at the end of the session, suggest that these students may have been evaluating the sensitivity of their decision rather than trying to force the system to "justify" a predetermined solution.

Student Survey

Descriptive statistics for student responses on the survey are given in Table 14. Detailed student responses to each question are given in Appendix G. In general, students reported a surprisingly low amount of knowledge of the risk-requirement-solution loop and expected value theory, given the survey was administered at the end of the course. The low values may also reflect the amount of instruction directly related to each topic (e.g., expected value theory is covered in a single lecture). Students also reported that the amount of time given for the knowledge mapping and Decision Analysis Tool tasks was adequate. With respect to students' perceptions about the knowledge map as an assessment-of-knowledge tool, students reported the knowledge map tool captured their understanding of the risk-requirement-loop somewhat, and that they were able to express their understanding in general using the tool.

Survey Results Concerning Decision Analysis Tool

Some of our most interesting findings emerged from the students' self-reports of impact of the Decision Analysis Tool. Overall, students reported some impact of the Decision Analysis Tool on their own understanding of the RAS problem, but much more impact on their team. Almost half of the students reported a moderate or large impact of the Decision Analysis Tool on their team's decision-making process and their team's discussion. Further, over half of the students reported the tool had at least some impact on their own understanding of expected value theory. Students found the tool useful and reported a willingness to use the tool in other parts of the course and they even suggested other Navy courses that might consider its use. The

finding that those students reported any impact at all is interesting given how little time students spent using the tool.

Correlations among the utility and impact of the Decision Analysis Tool are shown in Table 15. The pattern of significant correlations, together with the data in Table 14, suggests that the Decision Analysis Tool had its greatest impact on the team. That is, the tool may have been a focal point around which the team could jointly explore the problem space—the options, probabilities, and implications of different decisions are made explicit via the tool. The resulting discussion around these issues may be one of the biggest instructional outcomes.

Table 14

Student Survey Descriptive Statistics

Question	*N*	*Md*	*M*	*SD*	Min.	Max.
Please estimate how much you know about the risk-requirement-solution loop.[a]	17	3	2.35	1.17	1	4
Please estimate how much you know about expected value theory.[a]	15	3	2.33	0.98	1	4
How willing would you be to use the Decision Analysis Tool for other modules?[b]	17	3	3.12	1.11	1	5
Was the time given to you adequate to make your knowledge map?[c]	17	3	3.18	0.53	2	4
Was the time given to you adequate to use the Decision Analysis Tool?[c]	17	3	2.82	0.73	1	4
How well do you think knowledge mapping is able to capture your understanding of the risk-requirement-solution loop?[d]	17	2	2.00	0.61	1	3
How useful did you find the Decision Analysis Tool?[e]	17	2	2.47	0.80	1	4
In general, how difficult did you find the Knowledge Mapping Task with respect to being able to express your understanding of the risk-requirement-solution loop?[f]	17	2	1.88	0.86	1	4
What impact did the use of the Decision Analysis Tool have on your team's decision-making process (with respect to the RAS hiccup)?[g]	17	2	2.41	0.80	1	4
What impact did the use of the Decision Analysis Tool have on improving your team's discussion (with respect to the RAS hiccup)?[g]	17	2	2.29	0.69	1	4
What impact did the use of the Decision Analysis Tool have on uncovering gaps in your own understanding of expected value theory?[g]	17	2	1.76	0.97	1	4
What impact did the use of the Decision Analysis Tool have on improving your understanding of expected value theory?[g]	17	2	2.06	0.66	1	3

[a]1 = *Little or no knowledge*, 3 = *Somewhat knowledgeable*, 5 = *Very knowledgeable*. [b]1 = *Not willing at all*, 3 = *Somewhat willing*, 5 = *Very willing*. [c]1 = *Not enough time*, 3 = *About the right amount of time*, 5 = *Too much time*. [d]1 = *Not well at all*, 2 = *Somewhat well*, 3 = *Moderately well*, 4 = *Very well*. [e]1 = *Not useful*, 2 = *Somewhat useful*, 3 = *Moderately useful*, 4 = *Very useful*. [f]1 = *Not difficult*, 2 = *Somewhat difficult*, 3 = *Moderately difficult*, 4 = *Very difficult*. [g]1 = *Not much impact*, 2 = *Some impact*, 3 = *Moderate impact*, 4 = *Large impact*.

Table 15

Nonparametric (Spearman) Intercorrelations Among Decision Analysis Tool Perceived Utility and Impact Questions

	1	2	3	4
1. How useful did you find the Decision Analysis Tool?	–			
2. What impact did the use of the Decision Analysis Tool have on your team's decision-making process (with respect to the RAS hiccup)?	.67*	–		
3. What impact did the use of the Decision Analysis Tool have on improving your team's discussion (with respect to the RAS hiccup)?	.48*	.50*	–	
4. What impact did the use of the Decision Analysis Tool have on uncovering gaps in your own understanding of expected value theory?	.14	.23	.22	–
5. What impact did the use of the Decision Analysis Tool have on improving your understanding of expected value theory?	.09	-.16	.21	.41

*$p < .05$.

Discussion and Conclusion

We created computerized tools to support student assessment at the Navy's Engineering Duty Officer School at Port Hueneme, California. Rather than merely ask students what facts of knowledge they had acquired about the U.S. military's procurement process, we asked that they apply this knowledge in two ways: (a) use a knowledge mapping tool to diagram how risk management is incorporated into the U.S. military procurement process; and (b) use a decision analysis tool and an understanding of the concepts of expected value and multi-attribute utility theories to evaluate procurement options for a new RAS system.

In both cases, the context of the assessment was very different from other assessments that students had previously received. Unlike written tests, these assessments asked the students to diagram what they had learned or to apply what they had learned to resolve an authentic problem. Unlike the evaluation of student presentations, these assessments attempted to uncover the underlying process that students used to reach a conclusion. However, each assessment modality added to an ability to make inferences about the student's (or student group's) ability to apply what they had studied in their next duty assignments as Engineering Duty Officers.

While not statistically significant at the $\alpha = .05$ level, scores between pre- and post-test knowledge maps appeared to increase on average, and this trend becomes significant when $\alpha \approx .1$. We believe that a larger student cohort would generate a significant pre- and post-test difference at $\alpha = .05$. Considering that the problem-solving process was never explicitly taught in the course, but was presumably learned by students repeatedly applying the process in preparing for briefings throughout the short six-week course of instruction, we were encouraged by the positive trend in these results and the capability of the instrument to detect these changes.

Additional research is necessary to follow up on this result. The capability of an instrument to detect changes due to instruction is important validity evidence—the instrument should demonstrate sensitivity to changes in students' knowledge gains. In the current study, the hint at a difference is interesting because the students learned by doing rather than from didactic instruction. Nevertheless, we would require additional data to account for the part of this change that may have resulted merely from students becoming familiar with the HPKMT.

The Decision Analysis Tool (DAT) coupled with lag sequential analysis likewise provided insight on how the students incorporated risk management theories into the problem-solving process. In general, the students were less reticent to manipulate utility values than probability values and, in fact, they performed this type of action significantly more often than expected. The data collected from six student groups in two different classes suggested that all the students used one of either two different strategies:

1. Strategy 1: A global approach, in which students decide on all possible procurement options, and then set utility values and probability values independently of one another. Groups 1 and 4 displayed this type of strategy.
2. Strategy 2: An option-by-option approach where students set the utility values for each event of a single procurement option, then set the probabilities for the likelihood that each event within that option might occur. The process is repeated for each subsequent procurement option. In this strategy, the action of setting utility and probability values are more closely integrated than they are in Strategy 1. Groups 2, 3, 5 and 6 displayed this type of strategy.

While Strategy 2 was apparently the most popular way to approach the RAS problem, the amount of time students took to solve the problem did not seem to

predict a particular type of strategy usage. It is interesting to note, however, that the groups that used Strategy 1 were the groups that took the longest with the tool in their respective class group.

No matter which of the above strategies the students used in solving the RAS problem, however, lag sequential analysis was also able to detect other commonalities in the ways students used threshold and weighting events, as well.

The students often appeared to use threshold events to make the stoplights read a certain way. They would either respond to stoplight events by changing threshold values, or set threshold values to generate stoplight events. While this latter event could be interpreted as an attempt to check the sensitivity of their forecasts (Group 6), in most cases the data suggested the group was apparently using the tool to justify a predetermined course of action. For example, Group 1 finished their session by adjusting utility values and very quickly resetting the thresholds to generate previous stoplight settings. In light of this, we might conclude that the students generally viewed the threshold settings as a way to justify the choice of an option, rather than as an "absolute" measure of what exceeded minimum "caution" or "go" levels. While Groups 3 and 4 both set the upper threshold level prior to a stoplight event, no group fixed both upper and lower threshold levels as their first activity and left them unchanged until after they had fixed all the utility and probability values in the solution space. The one exception to this conclusion might have been the group with the seemingly most integrated solution (Group 5). Uncharacteristically from all the other groups, Group 5 never adjusted the threshold values in the problem space.

Similarly, most of the student groups also did not "weight" the relative importance of each of the multiple utilities in an event's overall utility until well into the problem solution. This suggests that students may either be uncomfortable with or unaware of (Groups 3 and 4) the idea of weighted utility values. Here again, Groups 1 and 5 set weights as one of their first activities and, interestingly, these groups spent the longest and shortest periods of time, respectively, using the tool. This suggests that the amount of time students spend with the tool probably does not significantly influence this activity.

The lack of a coherent strategy for setting threshold values and utility weights, as well as some groups' difficulty setting probability values, suggests to us that the students may be struggling to quantify these variables. In fact, both written

comments on the student survey as well as subsequent conversations with a number of students have confirmed this difficulty and suggest that, if this proficiency is an important aspect of EDO duties, the curriculum might address specific techniques to determine these values (e.g., economic forecasts, historical records, current priorities in the multi-year defense plan, etc.). Also since practitioners in the field of risk management suggest that, "the forecast is always wrong" (see DeNeufville, 1990, p. 273), it seems that EDOs might need to have this risk management skill.

While our analysis is able to distinguish various strategies, it cannot suggest that one strategy is "better" than another. As stated in our introduction, such an evaluation depends on the uses of our inferences. If students are being asked to mimic a specific procurement method, one strategy may always be preferred. On the other hand, if students must find support for a selected option, learning to amend values to rationalize a solution might be an important skill. Either way, we believe that the criteria against which strategies will be evaluated should be made clear to students if they are available. Similarly, evaluating the measurement quality of knowledge maps requires an evaluation referent. In addition to judging the effectiveness of strategies or knowledge maps, a referent would also allow us to identify those elements that define an "optimal" procurement strategy. For example, is it more important to consider the likelihood that future events will occur over the utility of those events? Are determining utility and probability (or other) attributes important in the skill set of EDOs?

Our experience with the six groups in this study suggests that if specific evaluative criteria could be established, the DAT tool could be developed into a tool that could both teach and evaluate student strategies in near-real time. As Wolf, Bixby, Glen and Gardner (1991) suggest, such assessments can become episodes of learning. To date, the development of the DAT has made use of the iRides simulation language and graphics, together with data-recording capabilities. The resulting product provides an environment in which concepts taught in the course can be applied in a tool context that can be quickly learned. However, the product does not make direct use of any of the iRides pedagogical features. As Self (1995) has shown, simply providing an interactive environment for experimentation is not sufficient to result in timely learning. BTL proposes to develop simple "How to" wizards using the LML lesson specification language. Students will be able to ask for quick reviews of basic concepts or assistance in carrying out steps in the use of the tool. The DAT data-recording scheme will be extended to detect instances of

students using these learning aids. In addition, we plan to offer a simplified DAT analysis of the restaurant case for use in the class on multi-attribute decision theory. This will make it possible for EDO School instructors to introduce core concepts of the tool in advance of teaching about Expected Value Theory. In light of the comments of Wolf et al. and Self, we believe that the HPKMT might also be more appropriately applied as a measure of how well the students understand how to quantify utility, probability, threshold, and weighting values contingent on an appropriate referent being developed.

References

Aguirre-Muñoz, Z. (2000). *The impact of language proficiency on complex performance assessments: Examining linguistic accommodation strategies for English language learners.* Unpublished doctoral dissertation. University of California, Los Angeles.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: an introduction to sequential analysis.* New York: Cambridge University Press.

Bakeman, R., & Quera, V. (1995). *Analyzing interaction: sequential analysis with SDIS and GSEQ.* New York: Cambridge University Press.

Baker, E. L., Linn, R. L., Herman, J. L., & Koretz, D. (2002). *Standards for educational accountability systems.* Retrieved March 22, 2004, from http://149.142.210.8/products/newsletters/polbrf54.pdf

Chung, G. K. W. K., Harmon, T. C., & Baker, E. L. (2001). The impact of a simulation-based learning design project on student learning. *IEEE Transactions on Education, 44,* 390–398.

Chung, G. K. W. K., O'Neil, H. F., Jr., & Herl, H. E. (1999). The use of computer-based collaborative knowledge mapping to measure team processes and team outcomes. *Computers in Human Behavior, 15,* 463–494.

Cole, M., John-Steiner, V., Scribner, S., & Souberman, E. (Eds.). (1978). *Mind in society: Development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Crouse, J., & Trusheim, D. (1988). *The case against the SAT.* Chicago, IL: University of Chicago Press.

deNeufville, R. (1990). *Applied system analysis: engineering planning and technology management.* New York: McGraw-Hill.

Educational Testing Service (n.d.) *Frequently asked questions about ETS.* Retrieved March 22, 2004, from http://www.ets.org/faq.html

Herl, H. E. (1995). *Construct validation of an approach to modeling cognitive structure of experts' and novices' U.S. history knowledge.* Unpublished doctoral dissertation, University of California, Los Angeles.

Herl, H. E., Baker, E. L., & Niemi, D. (1996). Construct validation of an approach to modeling cognitive structure of U.S. history knowledge. *Journal of Educational Research, 89,* 206-218.

Herl, H. E., O'Neil, H. F., Jr., Chung, G. K. W. K., & Schacter, J. (1999). Reliability and validity of a computer-based knowledge mapping system to measure content understanding. *Computers in Human Behavior, 15,* 315–334.

Klein, D. C. D., Chung, G. K. W. K., Osmundson, E., Herl, H. E., & O'Neil, H. F., Jr. (2002). *The validity of knowledge mapping as a measure of elementary students' scientific understanding* (CSE Tech. Rep. No. 557). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Lee, J. J. (2000). *The impact of Korean language accommodations on concept mapping tasks for Korean American English language learners.* Unpublished doctoral dissertation. University of California, Los Angeles.

Linn, R. L., Dunbar, S. B., Harnish, D. L., & Hastings, C. N. (1982). The validity of the Title I Evaluation and Reporting System. In E. R. House, S. Mathison, J. Pearsol, & H. Preskill (Eds.), *Evaluation studies review annual* (Vol. 7, pp. 427 – 442). Beverly Hills, CA: Sage Publications.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational Measurement* (pp. 13–103). New York: McMillan.

Munro, A. (2003). Authoring simulation-centered learning environments with Rides and Vivids. In T. Murray, S. Blessing, & S. Ainsworth (Eds.), *Authoring tools for advanced technology learning environments.* Dordrect, The Netherlands: Kluwer Academic Publishers.

Munro, A., & Pizzini, Q. A. (1998). *VIVIDS Reference Manual.* Los Angeles: Behavioral Technology Laboratories, University of Southern California.

Munro, A., Surmon, D., Johnson, M., Pizzini, Q., & Walker, J. (1999). An open architecture for simulation-centered tutors. In S. P. Lajoie & M. Vivet, *Artificial intelligence in education: Open learning environments: New computational technologies to support learning, exploration, and collaboration* (pp. 360-367). Amsterdam: IOS Press.

Munro, A., Surmon, D., & Pizzini, Q. (in press). Teaching procedural knowledge in distance learning environments. In H. F. O'Neil, Jr. & R. S. Perez (Eds.) *Web-based learning: Theory, research, and practice*. Mahwah, NJ: Lawrence Erlbaum Associates.

Norman, D.O. (1993). *Things that make us smart: Defending human attributes in the age of the machine.* New York: Addison-Wesley.

Osmundson, E., Chung, G. K. W. K., Herl, H. E., & Klein, D. C. D. (1999). *Concept mapping in the classroom: A tool for examining the development of students' conceptual understandings* (CSE Tech. Rep. No. 507). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Popham, J.W. (2000). *Modern educational measurement: practical guidelines for educational leaders.* Boston, MA: Allyn & Bacon.

Schacter, J., Herl, H. E., Chung, G. K. W. K., Dennis, R. A., & O'Neil, H. F., Jr. (1999). Computer-based performance assessments: A solution to the narrow measurement and reporting of problem-solving. *Computers in Human Behavior, 15,* 403–418.

Self, J. (1995). *Problems with unguided learning.* Proceedings of the International Conference on Computers in Education. ICCE'95.

Wiggins, G. P. (1993). *Assessing student performance: Exploring the purposes and limits of testing*. San Francisco: Jossey-Bass.

Wolf, D.P., Bixby, J., Glen, J., & Gardner, H. (1991). To use their minds well: investigating new forms of student assessment. *Review of Research in Education, 17*(2), 31-37.

Appendix A

Knowledge Mapping Task: Links Provided (Pilot Study)

EDO RAS Task
Online Knowledge Mapping Task
August 2003

Directions: Use the Knowledge Mapper software to depict your understanding of the steps of the decision analysis/risk management process for the RAS hiccup.

- Create a block diagram of the decision analysis process. Each block represents a step in the decision process.
- Use the reasons provided to show why you move from one step to another (e.g., Step A leads to Step B because of ______).
- Use the "*" if you think there is no explicit reason linking two steps (if one step logically and necessarily follows another).
- In addition to using the steps and reasons provided, you can type in your own steps and reasons if needed.

Notes:
- There is no one "correct" answer.
- You do not have to use all of the provided steps and reasons.

Steps in Process
analyze idea
consider impacts of 15 ships + parts
consider impacts of configuration management
consider impacts of contracting
consider impacts of late Phase A
consider impacts of manufacturing
consider impacts of R & D
consider impacts of supportability
consider impacts of test and evaluation
generate idea
prioritize cost drivers
prioritize decision drivers
prioritize performance drivers
prioritize schedule drivers
rate alternate ideas against environment with decision drivers
select option

Reasons for Next Step
*
decide
define the cost environment
define the environment
define the performance environment
define the schedule environment
score impacts

Appendix B

Knowledge Mapping Task: Links Not Provided (Pilot Study)

EDO RAS Task
Online Knowledge Mapping Task
August 2003

Directions: Use the Knowledge Mapper software to depict your understanding of the steps of the decision analysis/risk management process for the RAS hiccup.

- Create a block diagram of the decision analysis process. Each block represents a step in the decision process.
- Type in the reason why you move from one step to another (e.g., Step A leads to Step B because of ______).
- Use the "*" if you think there is no explicit reason linking two steps (if one step logically and necessarily follows another).
- In addition to using the steps provided, you can type in your own steps if needed.

Notes:
- There is no one "correct" answer.
- You do not have to use all of the provided steps.

Steps in Process
analyze idea
consider impacts of 15 ships + parts
consider impacts of configuration management
consider impacts of contracting
consider impacts of late Phase A
consider impacts of manufacturing
consider impacts of R & D
consider impacts of supportability
consider impacts of test and evaluation
generate idea
prioritize cost drivers
prioritize decision drivers
prioritize performance drivers
prioritize schedule drivers
rate alternate ideas against environment with decision drivers
select option

Reasons for Next Step
*
(type in your own reasons)

Appendix C

Knowledge Mapping Task (Main Study)

EDO Risk-Requirement-Solution Loop Task
Knowledge Mapping Task
November 2003

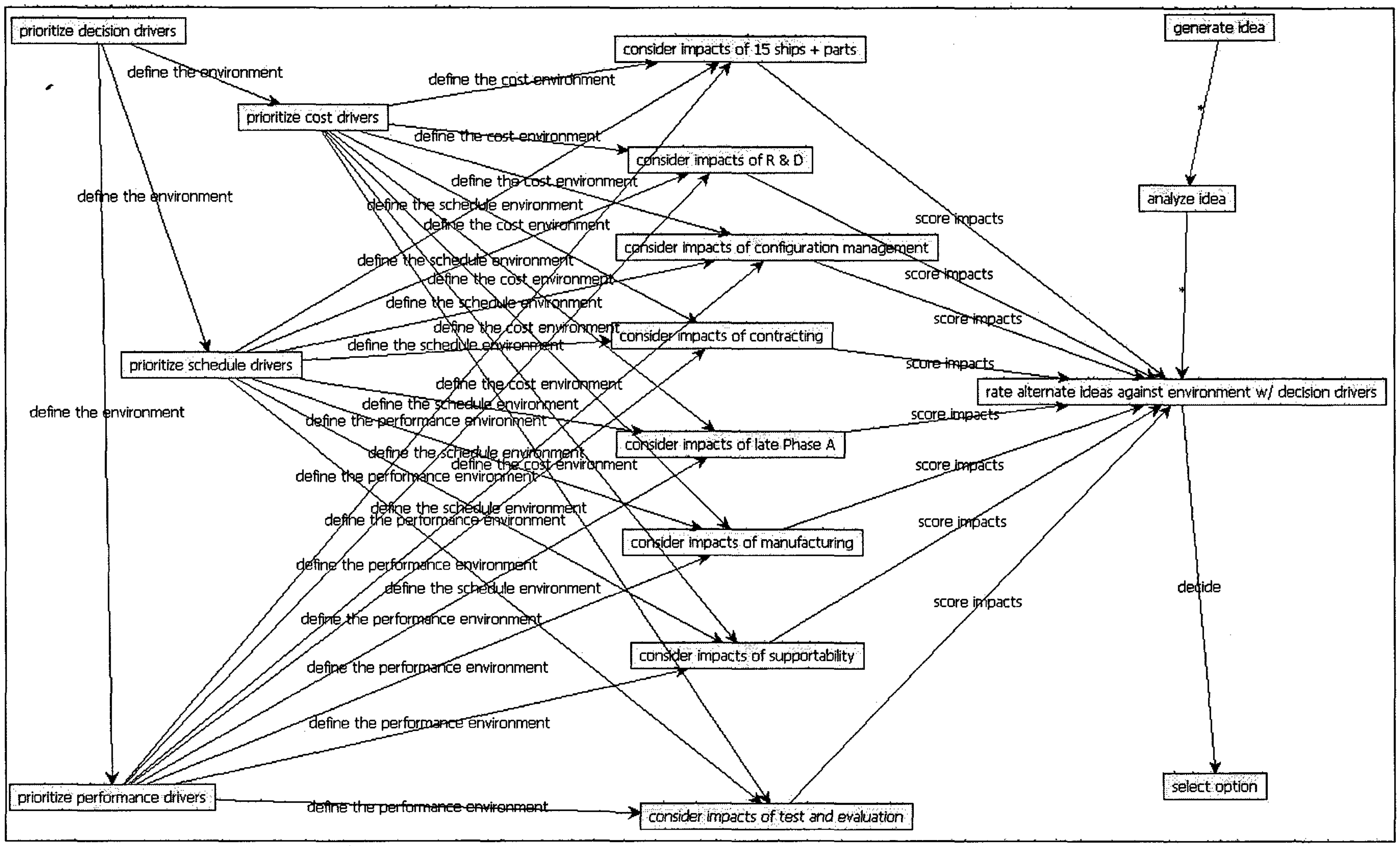
Directions: Create a knowledge map that depicts your understanding of the risk-requirement-solution loop using the steps provided below. Keep in mind the following:

1. There is no one correct answer.
2. Use as many or as few steps as needed. There are no irrelevant or "trick" steps. All steps are appropriate for this process.
3. If it's easier to use numbers instead of writing the entire label, please do so. Or abbreviate or use whatever method is easiest for you.
4. **Use each step only once**—cross out the step when you use it.
5. **Remember to put arrowheads** to show the direction between the steps.

Steps in Process
Conduct analysis of alternatives **1**
Define cost constraints **2**
Define legal constraints **3**
Define performance constraints **4**
Define requirement to address a risk **5**
Define schedule constraints **6**
Define the environment **7**
Detect new risks **8**
Execute best solution **9**
Methodically determine alternatives **10**
Methodically pick best solution **11**
Prioritize known risks **12**

Appendix D

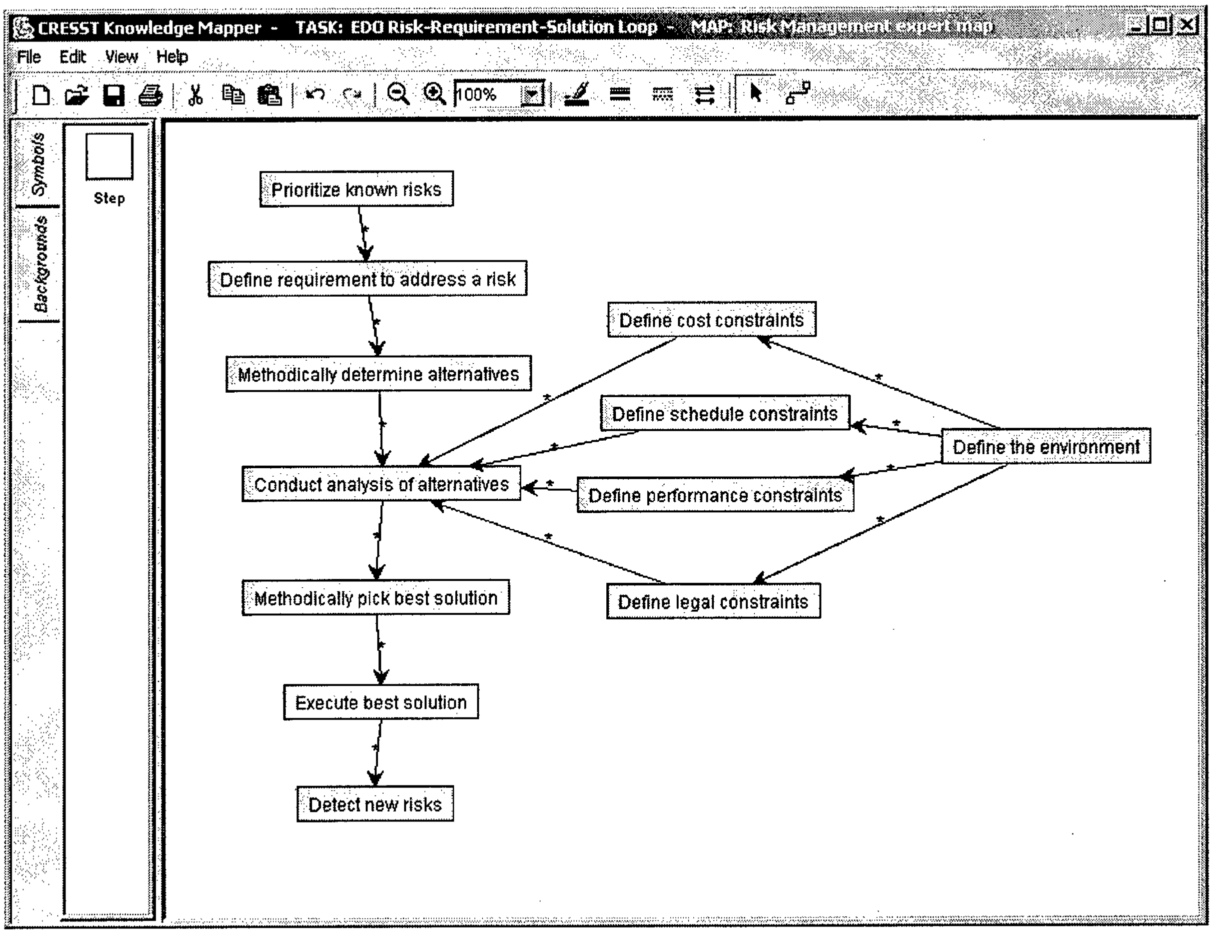
Criterion Knowledge Map (Pilot Study)



prioritize decision drivers
define the environment
prioritize cost drivers
prioritize schedule drivers
prioritize performance drivers
define the cost environment
define the schedule environment
define the performance environment
consider impacts of 15 ships + parts
consider impacts of R & D
consider impacts of configuration management
consider impacts of contracting
consider impacts of late Phase A
consider impacts of manufacturing
consider impacts of supportability
consider impacts of test and evaluation
score impacts
generate idea
analyze idea
*
rate alternate ideas against environment w/ decision drivers
decide
select option

Appendix E

Criterion Knowledge Map (Main Study)



Appendix F

Student Survey (Main Study)

EDO Study
Post-Task Survey
November 2003

1. Please estimate how much you know about the risk-requirement-solution loop.

❑ Little or no knowledge ❑ ❑ Somewhat knowledgeable ❑ ❑ Very knowledgeable

2. Please estimate how much you know about expected value theory.

❑ Little or no knowledge ❑ ❑ Somewhat knowledgeable ❑ ❑ Very knowledgeable

3. How willing would you be to use the Decision Aid Tool for other modules?

❑ Not willing at all ❑ ❑ Somewhat willing ❑ ❑ Very willing

4. Was the time given to you adequate to make your knowledge map?

❑ Not enough time ❑ ❑ About the right amount of time ❑ ❑ Too much time

5. Was the time given to you adequate to use the Decision Aid Tool?

❑ Not enough time ❑ ❑ About the right amount of time ❑ ❑ Too much time

6. When would be an appropriate time to introduce the knowledge mapping task?

❑ Beginning of course ❑ During program definition module ❑ During T & E planning module ❑ During acquisition logistics module ❑ During milestone review module

Other: ______________________________

7. When would be an appropriate time to introduce the Decision Aid Tool?

❑ Beginning of course ❑ During program definition module ❑ During T & E planning module ❑ During acquisition logistics module ❑ During milestone review module

Other: ______________________________

8. How well do you think knowledge mapping is able to capture your understanding of the risk-requirement-solution loop?

❑ Not well at all ❑ Somewhat well ❑ Moderately well ❑ Very well

9. How useful did you find the Decision Aid Tool?

❑ Not useful ❑ Somewhat useful ❑ Moderately useful ❑ Very useful

KNOWLEDGE MAPPING TASK QUESTIONS

In general, how **difficult** did you find the Knowledge Mapping Task with respect to…	Not difficult	Somewhat difficult	Moderately difficult	Very difficult
10. … being able to express your understanding of the risk-requirement-solution loop?	❑	❑	❑	❑

DECISION AID TOOL QUESTIONS

What **impact** did the use of the Decision Aid Tool have on …	Not much impact	Some impact	Moderate impact	Large impact
11. … your team's decision-making process (with respect to the RAS hiccup)?	❑	❑	❑	❑
12. … improving your team's discussion (with respect to the RAS hiccup)?	❑	❑	❑	❑
13. … uncovering gaps in your own understanding of expected value theory?	❑	❑	❑	❑
14. … improving your understanding of expected value theory?	❑	❑	❑	❑

15. Please write any other comments you may have about the Knowledge Mapping Task and the Decision Aid Tool below.

____

____

____

____

____

____

Appendix G

Student Survey Responses

	Little or no knowledge		Somewhat knowledgeable		Very knowledgeable
Please estimate how much you know about the risk-requirement-solution loop.	8	2	6	3	0
Please estimate how much you know about expected value theory.	4	3	8	2	0

	Not enough time		About the right amount of time		Too much time
Was the time given to you adequate to make your knowledge map?	0	1	14	4	0
Was the time given to you adequate to use the Decision Aid Tool?	1	4	12	2	0

	Beginning of course	During program definition module	During T & E planning module	During acquisitio n logistics module	During milestone review module
When would be an appropriate time to introduce the knowledge mapping task?	10	4	2	2	0
When would be an appropriate time to introduce the Decision Aid Tool?	1	9	6	3	2

	Not willing at all		Somewhat willing		Very willing
How willing would you be to use the Decision Aid Tool for other modules?	3	10	2	3	3

	Not well at all	Somewhat well	Moderately well	Very well
How well do you think knowledge mapping is able to capture your understanding of the risk-requirement-solution loop?	3	11	4	1

	Not useful	Somewhat useful	Moderately useful	Very useful
How useful did you find the Decision Aid Tool?	1	9	6	3

In general, how **difficult** did you find the Knowledge Mapping Task with respect to…	Not difficult	Somewhat difficult	Moderately difficult	Very difficult
… being able to express your understanding of the risk-requirement-solution loop?	7	8	3	1

What **impact** did the use of the Decision Aid Tool have on …	Not much impact	Some impact	Moderate impact	Large impact
… your team's decision-making process (with respect to the RAS hiccup)?	1	10	5	3
… improving your team's discussion (with respect to the RAS hiccup)?	1	11	6	1
… uncovering gaps in your own understanding of expected value theory?	8	9	0	2
… improving your understanding of expected value theory?	4	11	4	0

Additional comments
I'm curious about how each member had different ides on probability. As I understand it, the probability will be provided by data or proven performance. This leaves a lot of room for interpretation.
Need a little more time to fully understand the purpose and usefulness of the decision aid tool.
I see no other application for EDO Basic, but would be willing to use at my command.
Decision Aid tool was difficult in that it attempted to quantitize qualitative issues. Too easy to manipulative the desired answer in theory.
The tool as provided is insufficiently flexible to allow more than a very rudimentary analysis. It would have been easier to program my own analysis tool (using for instance, Matlab) if more sophisticated/flexible analysis were required
Either the knowledge map is not useful or explain in a vague way? Purpose?
Have a full lecture devoted only to learning the Tool, not just a 15 minute intro.
Design Aid tool is great and could be very useful. Expected value theory was never discussed.
Nice aid. The interface is a little counter intuitive, maybe a better "Windows" feel.
Decision Aid Tool would be useful in NPS-Naval postgraduate school sys eng and total ship systems eng curric's